1982

The Structure Of Associative Memory Traces: A Mathematical Analysis Of Learning Associative Clusters

Mark Lebas Howe

THE STRUCTURE OF ASSOCIATIVE MEMORY TRACES: A MATHEMATICAL ANALYSIS OF LEARNING ASSOCIATIVE CLUSTERS

by

Mark LeBas Howe

Department of Psychology

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
May, 1982

ABSTRACT

Two theories of associative memory that differ in the hypothesized structure of the memory trace (multiple independent traces or a single unitary trace) and in the way in which these representations are acquired (gradual strengthening or all-or-none) were examined in the context of a list-learning task. The associative, or multiple trace theory, characterizes memorial representation as a series of distinct memory locations that are independently linked together. Acquisition proceeds by continuously incrementing the strength of interconcept links. The Gestalt, or unitary trace theory, on the other hand, characterizes memorial representation as a single unitary structure. According to this model, acquisition consists of a discrete two-stage process. These theories were evaluated in two cued-recall experiments in which the degree of preexperimental knowledge was manipulated. Subjects learned lists of word triads (each of which consisted of a single cue and two target responses) to a stringent acquisition criterion. In Experiment 1, there were no preexperimental relationships between the members of the associative clusters. In Experiment 2, preexperimental knowledge was manipulated by varying the degree of intracluster category membership as measured by whether the cue and target items were typical or atypical category exemplars. In both experiments, a mathematical model that embodies stages-of-learning distinctions, was applied to the acquisition data. The results of these analyses indicated that: (a) Cues and targets were represented in a single holistic memory trace; (b) Learning consisted of a

discrete two-stage process; and (c) The manipulation of the degree of preexperimental knowledge affected the learning parameters of both acquisition stages, but had only a minimal impact on second-stage performance parameters. It was argued that these findings were consistent with a single unitary trace interpretation, namely, the modified storage-retrieval model.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Allan Paivio for his encouragement and guidance not only during the preparation of this thesis, but also throughout my graduate career. I am also grateful to Charles Brainerd who introduced me to mathematical modelling and offered much support during my development as a graduate student.

I would also like to thank Michael Pressley for his comments and suggestions on this research. A special note of thanks is offered to Julia O'Sullivan for sharing her support and for being so patient. I also appreciate the time and painstaking effort Lorrie Lefebvre expended in typing this thesis.

Finally, I would like to thank my parents, John and Margaret Howe, and my grandmother, Mabel LeBas for their support and encouragement through the years.

TABLE OF CONTENTS

Page

LIST OF TABLES

List of Tables (Cont'd)

LIST OF FIGURES

List of Figures (Cont'd)

LIST OF APPENDICES

INTRODUCTION

The general issue addressed in this paper is that of the nature of associative memory. Traditionally, this issue has been divided into two distinct but related questions: (1) What is the internal structure of associative memory representations, and (2) What is the nature of the processes involved in their acquisition? Essentially, the problem is this: When subjects memorize a list of items (e.g., words, sentences, etc.) where one or more components serve as a retrieval cue for the recall of the other components, how are the cues and targets represented in memory and what are the mechanisms responsible for their acquisition?

With respect to the first question, most memory theorists would agree that items in a list-learning situation may be conceptualized as bundles of potentially encodable features (e.g., Bower, 1967; Estes, 1959; Flexser & Tulving, 1978; Kintsch, 1974; Pelligrino & Salzberg, 1975; Underwood, 1969). However, there is little concensus on the way in which samples of these feature elements are internally represented. Similarly, with respect to the second question, there are few who would dispute the assumption that the basic memory processes involved in list acquisition, like those involved in most information-processing tasks, are executed in a chronologically ordered sequence of stages (e.g., Brainerd & Howe, 1982; Brainerd, Howe & Desrochers, 1980, 1982; Estes, 1978; Greeno, James, Da Polito & Polson, 1978). Indeed, it is not uncommon to think of items in a list "as passing through a sequence of stages with performance improving as items progress" (Brainerd, Desrochers & Howe, 1981, p. 2). While these stages have often been associated with familiar memory constructs such as encoding, storage, and retrieval, there is little agreement on how these processes operate.

The purpose of the present investigation was to evaluate two influential theories of associative memory. Both of these approaches, namely, the associationist or multiple trace theory (e.g., J. Anderson, 1976; J. Anderson & Bower, 1973; Ross & Bower, 1981; Underwood, Runquist & Schulz, 1959; Underwood & Schulz, 1960) and the gestalt or unitary trace theory (e.g., Flexser & Tulving, 1978; Greeno, 1970; Greeno et al., 1978; Halff, 1977; Humphreys & Greeno, 1970; Kohler, 1947), will be considered in detail below. Following this, both theories will be examined in the light of previous research. Most of this research has taken two distinct courses, one that corresponds to the question of item representation and the other that corresponds to the question of the nature of processes operating during acquisition. The data from both areas of research are found to be less than definitive with regard to the current controversy. In the former case, the interpretability of the data are limited by analytical as well as methodological constraints. In the latter case, while the data were found to support a unitary trace view, comparisons involved only a somewhat dated version of the multiple trace theory. More recent modifications of this position cannot be ruled out as (a) they were not directly tested, and (b) they are not entirely inconsistent with these data. Finally, it is difficult to obtain a comprehensive picture of the general nature of associative memory from these studies. This is because each area of research has dealt with a different aspect of this controversy and, in addition, there were methodological and analytical differences between these studies.

Next, two experiments designed to test a series of contrasting predictions derived from the more recent multiple and unitary trace positions will be outlined. A mathematical model that embodies stages-of-learning distinctions will also be presented and used to evaluate these predictions. This research represents an integrated approach to sorting out the general controversy surrounding the nature of associative memory. That is, both the question of item representation and the question of the nature of processes operating throughout acquisition are considered within the context of a single methodological and analytical framework.

In addition, these experiments involved a manipulation that not only served to further differentiate the multiple and unitary trace positions, but has also been the focus of recent theoretical interest, namely, preexperimental interitem semantic organization and its effects on cued recall. Specifically, the assumptions of these two theories were tested in conditions where preexperimental interitem organization (in this cse, taxonomic category membership) was absent (Experiment 1) and where it was present (Experiment 2). In Experiment 2, interest was also focused on the effects of variation in the degree or quality of interitem organization (by manipulating the typicality of category exemplars) on associative memory.

The Multiple Trace Theory of Memory

Historically, the multiple trace theory originated in classical philosophical associationism. This approach is both reductionistic and atomistic in that the whole of knowledge is reduced to a collection of constituent elements that are interconnected through direct mechanistic links or associative bonds. The fundamental claim of the associationist position (e.g., Underwood & Schulz, 1960) is that most, if not all, learning is rooted in the formation of these simple bonds between independent memory units. As such, the simplest version of the cued recall procedure, paired-associate learning, was considered the paradigm case of the elementary learning event. Here, subjects study a list of AB item pairs where the A term serves as the stimulus cue for the retrieval of the B target response. The major claim of the associationists is that the cues and targets are represented as distinct and independent memory traces that the subject has to link together.

The associationists posit three conditions, or processes, as necessary for learning. First, it is necessary to establish a representation of the cue which includes sufficient information to differentiate it from other cues (the process of stimulus cue differentiation). Second, it is necessary to establish a representation of the target which includes sufficient information to distinguish it from other potential targets as well as information specifying its production (the process of target response learning or integration). Third, it is necessary to establish an associative bond between the cue and the target that specifies which target is linked with each cue (the process of associative hook-up).

These different learning processes are said to be temporally ordered with respect to early and late phases in list acquisition. The logically and temporally prior stage in the learning process is the establishment of distinct cue and target traces in memory. As it is the target member of the pair which must be produced upon presentation of the cue, response learning is said to be the dominant process early in learning. Target acquisition, or integration, consists of developing a cohesive representation of the response, a process that is said to occur even when the responses are familiar and meaningful words. Once the cues have been differentiated, the targets integrated, and separate traces constructed for each, subsequent learning is said to consist of acquiring a direct trace-to-trace link connecting the cue and target. Once this hook-up stage is complete the subject is said to have learned the association.

More recent versions of the multiple trace theory have modified some of the features of this classical view. Current neoassociationistic models depict memory in terms of propositional networks (e.g., J. Anderson, 1976; J. Anderson & Bower, 1973; D. Norman & Rumelhart, 1975), frames (Minsky, 1975), schemas (e.g., Ross & Bower, 1981), and scripts (Schank & Abelson, 1977). All of these models represent items in memory as independent trace nodes rather than words. Like words, nodes are atomistic memory units; however, unlike words they are abstract entities that represent elementary ideas, meanings, or properties pertaining to the referent concept. Labeled, rather than simple unlabeled, associations connect these idea nodes by expressing propositionalized relations between concepts in a manner similar to the case relations described by Fillmore (1968). It has been argued that the advantage of substituting nodes for words and labeled for unlabeled links is that it provides a more flexible system for representing knowledge. This flexibility enables the theorist to construct and express more complex, variable, and numerous concepts and relations than was possible in classical associationism. It should be pointed out, however, that despite a strong theoretical committtment to the idea that abstract nodes become connected through elaborate propositional tree structures, many investigators proceed experimentally as if words were connected by single unlabeled links, much in the same manner as their predecessors.[1]

The framework developed in the traditional associationistic analysis of learning is still generally accepted in these more recent versions of the multiple trace theory. Items in a list-learning experiment are represented as independent memory units that must be linked together. Encoding, storage, association, and retrieval all proceed under the assumption that both the individual nodes and their connecting pathways are established with independent probabilities. Unlike traditional associationism, however, storage of item information plays a relatively minor part in the acquisition process as conceived in neoassociationism. While it is generally acknowledged that node activation precedes the acquisition of a cue-target association, it is not clear that this activation constitutes a separate, temporally prior stage in the learning process. Indeed, cue and target storage (i.e., stimulus and response learning) which had been considered a critical component in list acquisition, <u>even when the material was familiar and meaningful</u>, is now all but ignored. In fact, the current position is that, "If the stimuli and responses are common words then (the learning system) does not have to bother learning them since they are already represented by word nodes in memory" (J. Anderson & Bower, 1973, p. 423).[2]

Rather than viewing the processes of information storage and cue-target association as distinct and equally important operations, recent versions of the multiple trace position consider the associative process to be the key determinant in item acquisition. This is the crux of the two versions of the multiple trace theory that are considered in this investigation, namely, J. Anderson and Bower's Human Associative Memory (HAM) model and Ross and Bower's (1981) schema model. In both models the critical variable affecting associative operations, and therefore learning itself, is the availability of, or ease of constructing, interitem propositions.

In HAM acquisition consists of locating, activating, and tagging (usually on the first study trial) the concept nodes in memory that correspond to the words presented in the study list. This activation then spreads outward from the cue and target nodes in "search" of intersecting pathways. If an intersection, or propositional relation, is found between these concepts it is primed and tagged for later use. If no relationship is found then the learner must construct an "artificial" proposition such as, "<u>A</u> is paired with <u>B</u>" which connects the two terms. This latter process is more demanding, making learning more difficult, than simply locating and tagging a preexisting pathway. At retrieval, the "executive" examines the various associative branches radiating out from the search cue node to check whether any of them have been marked with a list tag as a retrieval route to another list item. If a tagged proposition is found it is followed to its terminus whose node may similarly be marked with a list tag, in which case it will be recalled.

The rate of learning in this model is said to be determined by whether or not a preexisting relation is available between the cue and target nodes. This is an all-or-none affair in that there either is or is not a preexisting relation, and does not depend on the degree (strong or weak) of relationship between the items. That is, the "... rate of 'associative learning' of an S-R pair will not be dependent on where the response occurs in the rank-ordering of the associate in the norms for the S-term; rather, learning rate will depend on whether or not any preexperimental association is available from the S-concept to the R-concept" (J. Anderson & Bower, 1973, p. 425).

Insert Figure 1 about here

Ross and Bower's (1981) schema model (shown in Figure 1) is viewed as a simplified version of the more general multiple trace position and was developed specifically for list-learning experiments. While it has many features in common with HAM, the rate of associative learning is not only determined by the availability of a preexisting relation, but also by the degree of relatedness between the concepts. In order to represent degrees of relatedness, a strength index has been added to each memory proposition. Each link emanating from a node has a certain initial strength value associated with it. Every time a learner receives additional experience with the proposition (e.g., a study trial or a successful retrieval attempt on a test trial), the strength of that proposition is incremented. In this version of the multiple trace theory then, learning consists of, "... strenghtening the preexisting associations among the presented items" (Ross & Bower, 1981, p. 5).

In this model, only indirect associations are permitted among concepts; that is, associations between items are mediated through a schema or grouping node which is conceived of as a proposition or set

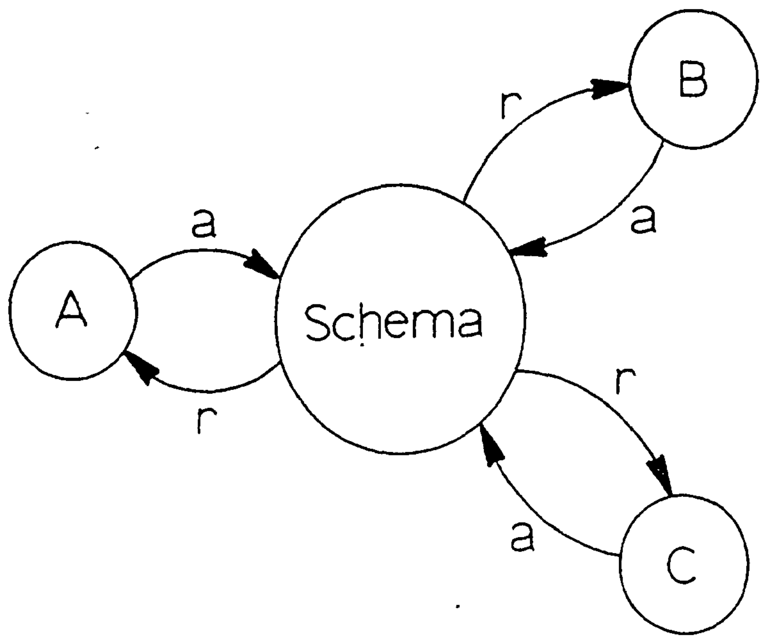

Figure 1. A Schematic Rendering of the Hypothetical Structure of a One Cue (A) - Two Response (B and C) Associative Memory Representation According to the Multiple Trace Model. (Adapted from Ross & Bower, 1981, p. 3.)

of propositional elaborations. When items are presented for study, the concept nodes to which the words refer are activated in memory. This activation spreads in a parallel limited-capacity manner along pathways that branch from each node, the result of which may be the discovery of an intersection between two or more of the concepts. If an intersection, or in this case a schema, is found the initial strength of pathways leading to the schema, as well as those leading back to the concepts, is incremented.

This model is characterized by two parameters _a_ and _r_. The parameter _a_ is an "access parameter" that measures the likelihood that a retrieval cue accesses the memory schema. The parameter _r_ is a "response parameter" that measures the likelihood that a target response will be recalled given that the schema has been activated by the retrieval cue. Learning not only consists of tagging the specific components of the schema that are members of the study list, but also gradually incrementing the strength value of the cue-to-schema link (as reflected in changes in the _a_ parameter) and the schema-to-target links (as reflected in changes in the _r_ parameter). At retrieval, the schema must be activated first as access to the responses is mediated solely through this node. Once the schema is activated, recall of the target responses will occur given that the pathways leading from the schema to the responses are of "sufficient" strength. Implicit in this model is the assumption that the parameters _a_ and _r_ must exceed some critical strength value, or threshold, as retrieval of each response is viewed as an all-or-none process (i.e., an item either is or is not recalled). Indeed, all associations are said to be established independently in an all-or-none manner.

The Unitary Trace Theory of Memory

The origin of the unitary trace theory is generally associated with the gestalt school of psychology (e.g., Asch, 1969; Koffka, 1935; Kohler, 1940, 1941, 1947). (A similar view was espoused nearly a decade earlier by Semon whose work has recently been reviewed by Schacter, Eich and Tulving, 1978.) In contrast to the reductionistic and atomistic view of the associationists, this approach views memory units as complex wholes that are not simply the sum of their individual components. The essence of this approach is that what is stored in memory is a single unified event. In sharp contrast to the multiple trace theory, gestalt psychologists argued that information about both the cue and the target were represented as a single, unitary trace in memory.

Considerable emphasis is placed on organizational processes. Indeed, the very process of association is a form of cognitive organization where subjects must bring both the cue and the target into a single relational structure. In fact, " ... association depends upon organization, because an association is the after-effect of an organized process" (Kohler, 1947, p. 163).

While organization is also an important component of the multiple trace theory, its role is conceptualized in a framework that is entirely different than the one proposed in the unitary trace model. From an associative point of view, organization is that process whereby collections of distinct concepts become linked to one another, but each concept still maintains its original independent identity. That is,

The sole formal content of the concept association is

> that of a 'connection' or 'path' that does not alter the terms connected. The only type of interdependence that associative accounts admit is therefore that of links in a chain; the terms connected are one thing, the connection between them another.
>
> (Asch, 1969, p. 97)

Contrariwise, the unitary trace theory, being more holistic in orientation, says that the process of organizing concepts into a single representational unit involves more than simple connecting otherwise distinct concepts. Rather, the concepts and their relation are combined to form a single integral term in memory. In other words,

> When an A and a B become associated, they are experienced not as two independent things but as a member of an organized group-unit ... (which) ... cannot consist of two separate parts of which one corresponds to A and the other to B. Rather, the unitary experience indicates that a functional unit is formed ... in which the processes A and B have only relative independence, ... (therefore) we cannot expect two separate traces ... only one trace will be established, which represents the functional unit ... .
>
> (Kohler, 1947, pp. 159-160)

Organization not only plays a key role in the representation of information in memory, but is also at the core of the learning process itself. Early learning not only consists of forming a single relational structure in which information about the cue and the target is integrated in memory, but also involves learning to recognize the items as ones for which a unitary trace has been stored. Later acquisition consists of learning to locate the trace reliably given the stimulus as a retrieval cue. This interpretation is consistent with Asch's (1969) position where subjects must first learn to recognize what information has been stored in the unitary memory representation, and then must learn how to recall it. That relational storage is followed by learning to retrieve is also consistent with evidence that learning to recognize is followed by learning to recall (Estes & Da Polito, 1967; Kintsch & Morris, 1965).

More recent versions of the unitary trace model include storage-retrieval theory (Greeno, 1970; Greeno et al., 1978; Halff, 1977; Humphreys & Greeno, 1970; Pagel, 1973) and encoding-specificity theory (Flexser & Tulving, 1978; Tulving, 1979). As in the original gestalt theory, the process of forming an association consists of representing the cue and target components in an integrative cognitive unit; that is, a relational structure that incorporates the associated elements in an integral fashion. The internal representation of this unitary trace can be conceived of (as it is in encoding-specificity theory) as a subset of N features of potentially encodable elements from the presentation of the cue and target items. These traces, one for each cue-target combination, are conceptualized as single vectors, or linear arrays, containing the N elements or dimensions (see Figure 2). Feature encoding is equivalent to assigning values to the dimensions that represent the various features in the memory trace. (The values assigned to the various dimensions are assumed to vary in a continuous fashion. While this is the preferred representation, Flexser and Tulving acknowledge that an equivalent model could be constructed on the basis of binary feature representation. For the present purposes, either representation is acceptable.) Over a sample of study events the number of encoded features will vary in a probabilistic manner, with unencoded features having undefined values.

In the storage-retrieval theory, acquisition is broken down into a sequence of chronologically ordered events. These events constitute two discrete all-or-none stages. The first stage consists of storing a unitary trace and learning to recognize it in memory. The second stage consists of developing a reliable retrieval plan. As the first stage consists of storing an integrated cue-target representation, the difficulty of early learning is determined jointly by stimulus factors, response factors, and the number of features they have in common. The greater the number of shared features the easier it will be to construct a unitary trace.

Insert Figure 2 about here

There are two interpretations of events that occur in the second stage of learning. In the first of these, both recognition and recall are viewed as examples of attempts to retrieve the memory trace by matching the features of the cue to those in the unitary trace. In storage-retrieval theory, the difficulty of this second stage depends to a lesser extent on target manipulations than on cue manipulations as acquisition of a retrieval system is said to consist of developing a network of stimulus feature tests. In encoding-specificity theory

FEATURE

	1	2	3	4	5	6	7	8	9	10	11	12
UNITARY MEMORY TRACE	1.9		6.3	2.8	4.5			3.6	5.8		7.9	2.0
RETRIEVAL CUE		2.2	6.3		4.5	9.7	1.4		5.8		7.9	

Figure 2. A Schematic Rendering of the Hypothetical Structure of a One Cue - Two Response Associative Memory Representation According to the Unitary Trace Model. (Adapted from Flexser & Tulving, 1978, p. 161.)

the important component in recall is not only that the retrieval cue be a part of that which was originally stored in the trace, but also the degree of correspondence between the cue and the unitary representation. As recall attempts are viewed as a process whereby features of the cue are matched to those in the unitary trace, the difficulty of later learning depends more on cue than target manipulations. The matching process operates in an all-or-none fashion. That is, even though the features may be represented in terms of continuous values, the features of the stored trace either do or do not match those of the retrieval cue. The usefulness of a feature at the time of retrieval depends, therefore, not only on its presence in both the stored traced and the current encoding of the retrieval cue, but also upon the degree of correspondence between the value of the feature in memory and its value in the retrieval cue. The more features that are shared between the cue and the trace, the easier retrieval learning is.

A second interpretation of retrieval learning has been presented by Halff (1977) and Brainerd et al. (1980, 1981). In this extended storage-retrieval version there are two types of retrieval systems, namely, heuristic systems and algorithmic systems. Heuristic retrieval corresponds to a general plan that is available immediately for recall of newly learned material and is essentially independent of the nature of the stored material (i.e., context-free operations). One disadvantage of this kind of retrieval operation is that it is error prone and therefore does not permit perfect retrieval of the material that subjects are required to learn. Algorithmic retrieval, on the other hand, consists of specific plans (i.e., context-sensitive operations) whose advantages include near perfect recall. Algorithmic retrieval has two components, " ... namely, a search operation that finds the item in long-term memory and a response decoding operation that is activated when the item is found" (Brainerd et al., 1981, p. 13). As the former operation depends primarily on factors related to the retrieval cue and the latter operation depends primarily on factors related to the target items, both cue and target manipulations are important in second-stage, as well as first-stage learning.

As the fundamental learning event in the unitary trace theory is the organization of information into a single integral memory unit, the presence as well as the degree of preexperimental interitem organization is important in both learning stages. Increasing the amount of interitem organization should facilitate storage of a unitary trace in that as the number of common features increases the greater the likelihood is of discovering a relational base mediating the association. Similarly, at retrieval, increasing the number of common features should also increase the probability of a match between the cue and the trace, thereby facilitating search and decoding operations.

Empirical Comparison of the Multiple and Unitary Trace Models

There have been two independent lines of research that have addressed different aspects of this controversy surrounding the nature of associative memory. The first area of research has focused primarily on the question of whether items are represented holistically or independently in memory. The second area of research has concentrated more on an examination of the temporal course of the processes involved in acquisition. Both of these areas are discussed below.

The question of whether cues and targets are represented in a single integral unit or are represented in separate traces that must be linked together, has been investigated in the context of sentence memory research. While the topic of sentence memory was never directly addressed by gestalt writers, various authors (e.g., J. Anderson, 1976; J. Anderson & Bower, 1971, 1972, 1973; Goetz, R. Anderson & Schallert, 1981) have conjectured that a unitary trace model would assume that lingustic ideas are integrated into holistic representations. As such, these representations are configural in nature, having the getalt property of emergent information. That is, the configural representation must include properties that are different from and go beyond the simple sum of the linguistic associations between the individual sentence elements. On the other hand, neoassociationistic multiple trace theories (e.g., J. Anderson, 1976; J. Anderson & Bower, 1971, 1972, 1973) propose that the associative structure underlying sentence memory does not contain holistic representations with emergent properties. Instead, as each association is stored independently, sentence memory is nothing more than the sum of the individual associations between sentence elements.

In order to test these alternatives, Bransford and Franks (1971) presented subjects with a list of sentences that expressed a series of simple ideas. Memory was assessed using a recognition test that included the original sentences as well as sentences that expressed logical, but novel, combinations of the original simple ideas. In addition to correctly recognizing the old sentences, subjects consistently and just as confidently, incorrectly recognized the novel sentences which expressed complex combinations of the original ideas. These results have been obtained for both concrete and abstract material (Franks & Bransford, 1972), and suggest that subjects form integrated ideas which have emergent properties that are different from the simple aggregate of independent linguistic associations.

Evidence for holistic, integrated sentence representation has also been obtained under conditions where sentence retrieval was cued with words that, while not presented at study, were judged close in meaning to the sentence as a whole. For example, R. Anderson and Ortony (1975) presented subjects with a list of sentences and cued recall with a word which was close in meaning to either the sentence as a whole, the subject of the sentence, or the sentence predicate. If the multiple trace view is correct, and the constituents of the sentence are represented independently, then the probability of recall when presented the holistic retreival cue should equal the sum of the probabilities of recall when presented with the subject cue alone and the predicate cue alone. Alternately, if the unitary, configural trace view is correct, and the whole is truly greater than the sum of the independent associations, then the probability of recall given the holistic retreival cue should exceed the sum of the recall probabilities of the component cues. This latter result was consistently obtained, and has been shown to hold for both concrete and abstract material (Marschark & Paivio, 1977).

Evidence favoring the associative multiple trace view has been obtained in a number of experiments reported by Anderson and his associates (J. Anderson, 1976; J. Anderson & Bower, 1972, 1973). J.

Anderson and Bower devised a "cross-over" sentence memory experiment. Subjects learned lists of sentence pairs which contained the same object but had different subjects and verbs. Retrieval of the object term was cued with either the subject or verb alone (one cue conditions), the subject and the verb from the same sentence (congruent two cue condition), or the subject and verb from different sentences (cross-over two cue condition). As most theories would have accurately predicted that the congruent two cue recall condition would be superior to either of the one cue conditions, the critical comparison was between the congruent and cross-over conditions. As each sentence is represented as a single unitary trace, the configural theory should predict that recall in the congruent cue condition should exceed recall in the cross-over condition. Counterintuitively, the multiple trace theory predicts the reverse. As the cues in the cross-over condition represent two independent routes to the object, while the cues in the congruent condition represent only a single access route, there should be evidence of additive retrieval effects in the former but not the latter condition. In accord with the associative view, additive effects were consistently obtained.

Subsequent research, however, has not always replicated these results. For example, Foss and Harwood (1975) obtained nonadditivity. In addition, they found that the associative model failed to pass a much simpler test (Foss & Harwood, 1975, Experiment 2). Multiple trace theories, which assume independence of associations, make the simple prediction that the probability of recalling the object of a sentence given both the subject and verb as cues, should be equal to the sum of the probabilities of recalling the object given either the subject or the verb as cues. In contrast, the results consistently showed that the probability of recall in the two cue condition was significantly higher than the sum of the probabilities for the one cue conditions.

In the research reviewed above, the controversy of whether the memory trace of a sentence is configural or associative has been reduced to the question of whether two or more retrieval cues act independently or combine interactively. The independence position says that multi-cue retrieval is predictable from the sum of the independent operation of each cue, whereas the configural position predicts that multi-cue retrieval is more than the simple additive combination of the independent cues. The evidence reviewed above provides only equivocal support for either of these positions. In fact, there has not been a clear and consistent demonstration to date that cues act either completely independently or in a completely nonadditive fashion. A particularly lucid example of this point can be found in J. Anderson's (1976, pp. 415-417) summary of nine sentence memory experiments. Here, the degree of independence of items in memory was affected simply by whether or not subjects were instructed to meaningfully process the sentences during encoding. That is, nonindependence was obtained when subjects were instructed to process the sentence in a meaningful way, but not when subjects were given standard study instructions.

Hintzman (1980) views this as a perfect illustration of Simpson's paradox. Typically, tests of stochastic independence involve the construction of 2 x 2 contingency tables that relate success and failure on one retrieval attempt to success or failure on another attempt. Simpson's paradox occurs when two or more of these contingency tables are combined (e.g., by collapsing across subjects and items). This new summary table may show relationships that are different from those in the original data, particularly when one or more hidden variables are correlated with the original variables of interest, or are correlated with the relationship between them. Such summary analyses can, therefore, be misleading as the apparent strength of the relationship between the original variables can increase or decrease as a function of this hidden variable. Even more problematic is the possibility that the relationship could reverse direction or that a spurious relationship, even spurious independence, may result.

Hintzman (1980, pp. 404-406) argued that in the case of sentence memory experiments, particularly those reviewed by J. Anderson (1976), spurious stochastic independence could have been obtained using such summary data, in spite of the fact that the data may have been more consistent with a configural hypothesis. Instructions to process the sentences meaningfully may have induced subjects to attend about equally to each sentence.

Alternately, subjects who received no orienting task may not have distributed their attention equally to all sentences. One effect of the former manipulation may have been to reduce subject-item encoding interactions which normally disguise the configural result of nonindependence. In the latter condition, there is a greater likelihood of obtaining spurious independence, as the data are summarized across potentially substantial differences in subject-item encoding operations. Given this line of reasoning, data which exhibit nonindependence under standard study conditions (e.g., Foss & Harwood, 1975; Jones, 1976; 1978b) make an even stronger case for the configural hypothesis. This is because the direction of the deviance from independence is the opposite of that expected from subject differences, item differences, and subject-item interactions.

Regardless of whether or not Simpson's paradox is viewed as problematic in this area of research (see Martin, 1981, for arguments against Hintzman's claim), these data are subject to another criticism that has recently been raised in connection with the general area of cued recall. This criticism is methodological in nature and concerns the fact that each of these studies is limited by the use of a fixed-trials design. That is, in all of these experiments only a small number of study-test cycles are administered (usually less than six) on relatively long lists (usually consisting of twenty to thirty items). In such cases it is not at all clear that learning has been equated across subjects and items. Given the general assumption, outlined earlier in both the multiple and unitary trace theories, that memory operations are executed in a sequence of stages, experiments such as these face two problems: "(a) The data will invariably underestimate the impact of manipulations that primarily affect later learning relative to the impact of manipulations that primarily affect early learning, and (b) list difficulty comparisons are confounded with completeness of learning, and this can produce inconsistent list-difficulty orderings when experiments vary on methodological factors (e.g., list length, number of study-test cycles) that affect completeness of learning" (Brainerd et al., 1981, p. 2).

Both of these points pose interpretive dilemmas for cued recall experiments in which the effects of memory variables (e.g., organization, instructional set) are examined. For example, with respect to Point a, the best that can be expected using a fixed-trials design is the completion of only the very earliest of the learning stages as this procedure does not insure that all items pass through the entire acquisition sequence. What this means is that the difficulty of early learning stages will contribute more to global statistics (e.g., total errors) than the difficulty of later stages. Therefore, the effects of variables that affect the ease of later learning but not early learning, will be underestimated and sometimes even dismissed. Similarly, when two treatments are being compared, where one primarily affects early learning and the other later learning, the former manipulation will be judged to have a larger effect on performance than the latter manipulation, " ... regardless of the true magnitude of the treatments' effects on their respective stages" (Brainerd & Howe, 1982, p. 4). Even more problematic is that the degree of underestimation increases as the number of study-test cycles decreases and is maximal when only a single cycle is administered, a procedure which typifies many of the sentence memory experiments discussed above.

The important aspect of Point b is that because fixed-trials designs limit the number of acquisition stages that items pass through, treatment effects are confounded with completeness of learning. That is, " ... if the same number of study-test cycles is administered with lists of different difficulty, amount of learning, in the sense of how many stages have been completed, will usually be greater for items on easier lists" (Brainerd et al., 1981, p. 2). What this means is that the difficulty of both early and late stages in the acquisition sequence will contribute to the performance measure for easier items, whereas only early learning will be measured for harder items. This problem is of particular concern when comparisons are conducted between different experimental conditions, or between experimental and control conditions, when the manipulations have asymmetrical effects on acquisition stages, or their effects on different stages are unknown (see Brainerd et al., 1981 and Brainerd & Howe, 1982, for specific examples of this problem).

This latter point is particularly relevant when considering the seemingly contradictory results of the experiments reported by J. Anderson (1976). It is well known that instructing subjects to process items meaningfully at the time of study improves subsequent recall over standard control conditions (e.g., Craik & Lockhart, 1972). Even though it is not known whether variation in instructional sets has asymmetrical effects on different learning stages, it is clear that items in this condition will have been more completely learned than those in the control condition. It may be, then, that when level of learning between conditions is equated, the data will be more consistent with the configural or unitary trace view than with the associative or multiple trace view.

The second area of research to be considered here, one that is concerned with examining the temporal sequence of acquisition processes, deals directly with the issue of completeness of learning. In these experiments, Greeno and his associates (e.g., Greeno, 1970; Greeno et al., 1978; Humphreys & Greeno, 1970) tested the different multiple and unitary trace predictions about processing stages in list acquisition. In order to equate levels of learning in these studies, subjects were required to meet a stringent acquisition criterion (e.g., two consecutive passes through the list without an error). Using this procedure, equal weight was given to both early and late stages in acquisition and item difficulty was not confounded with completeness of learning.

In a lengthy series of PAL studies (reviewed in Greeno et al., 1978) Greeno tested his storage-retrieval theory and the traditional associationistic account of learning. It will be recalled that in this latter view it is proposed that early acquisition consists primarily of response learning and that later learning consists of cue-target hook-up. In storage-retrieval theory, on the other hand, it is proposed that early learning consists of storing a permanent unitary encoding of the cue and the target whereas later learning consists of developing an infallible retrieval plan. These two hypotheses can be rephrased in terms of the question of whether early learning focuses primarily on the encoding and storage of response information, or whether information about both the cue and the target is involved in the learning process from the outset.

In order to test these different predictions, Greeno used a mathematical model that embodies stages-of-learning distinctions. In terms of this two-stage learning analysis, the multiple and unitary trace interpretations lead to different predictions about the behavior of parameters that measure learning in these stages. As response learning is said to precede learning a connection, strong support for the associationistic model would consist of finding that the manipulation of response difficulty affected the parameters associated with the first stage, that second stage parameters were affected by both cue and target difficulty manipulations, and that the magnitude of the cue and target effects were equivalent in the second stage. Somewhat weaker support would consist of demonstrating that response effects tend to decrease or that stimulus effects tend to increase as learning proceeds from early to late stages. The storage-retrieval theory, on the other hand, makes much simpler predictions. If acquisition consists of learning to store and recognize followed by learning to retrieve a unitary relational structure, then cue and target difficulty manipulations should affect parameters of both stages in a roughly equivalent manner.

In general, the bulk of the evidence from the studies reviewed by Greeno et al., which included manipulations of meaningfulness, familiarity, and concreteness, among others, was in agreement with the predictions from storage-retrieval theory. For example, Humphreys and Greeno (1970) and Greeno and Marsh (cited in Greeno et al., 1978) found that first-stage difficulty was affected by both cue and target familiarity manipulations. Greeno, Hoffman, Shaffron and Menuhin (cited in Greeno et al., 1978) found similar effects when meaningfulness was manipulated. Curiously, however, second-stage parameters were affected only by stimulus manipulations in the familiarity experiments and interacted with the cue and target when meaningfulness was manipulated. For example, stimulus difficulty affected both stages when the responses were hard, but only the initial stage when the responses were easy. Similarly, response difficulty affected both stages when the stimuli were hard, but only the initial stage when the stimuli were easy. While Greeno interpreted these results as suggesting that learning to retrieve was primarily a problem of stimulus discrimination, Brainerd, et al., (1980) suggested that these anomalous findings may have been due in part to the fact that cue and target difficulty manipulations were asymmetrical (i.e., that the difficulty manipulation was harder on the stimulus side than on the response side for some lists and vice versa for other lists). When symmetrical familiarity and concreteness manipulations were used, Brainerd et al. found that learning parameters in both stages reacted in a qualitatively similar manner to cue and target difficulty. While other differences between these studies may have contributed to the different outcomes (e.g., variations in list presentation procedures), the data converge on the conclusion that learning operates on unitary rather than multiple traces in the traditional associationistic sense.

THE PRESENT RESEARCH

The data from the research reviewed above are equivocal with regard to the controversy surrounding the nature of associative memory. The sentence memory research considered only a limited aspect of this controversy, namely, the question of item independence in memory, and was found wanting on both statistical and methodological grounds. That is, these data were limited in their interpretability due to the potential analytical problems associated with Simpson's paradox as well as the problems inherent in fixed-trials designs. The research by Greeno et al. and Brainerd et al., while eliminating the fixed-trials confounds, focused exclusively on the chronological sequence of acquisition process and did not directly test the assumption of item independence. Although the data provided strong support for the unitary trace position, this research does not rule out the possibility that cues and targets are represented independently in memory as only one (and perhaps a somewhat dated) version of the general multiple trace position was tested. And, as Greeno et al. point out, even this traditional associative view can be modified to account for the main trend of the data in a manner consistent with more recent versions of the multiple trace theory. For example, the mere fact that both cue and target difficulty affected early learning does not rule out the possibility that items are independently represented in memory as it could be assumed that the first stage of learing consists of developing separate stimulus and response representations. Similarly, the fact that both cue and target difficulty affected later learning is not inconsistent with the assumption that second-stage learning consists of finding a connection between the separate stimulus and response traces.

The purpose of the present investigaton was to integrate these separate lines of research into a single unified test of both the structural and processing assumptions associated with the more recent multiple trace theory and the unitary trace theory. Specifically, a mathematical model that embodies stages-of-learning was used to test: (1) the multiple trace assumption that learning consists of establishing independent associative links between separate memory traces and the unitary trace assumption that acquisition involves the establishment of a single trace in which item learning resembles a dependent process, and (2) the multiple trace assumption that learning over a series of trials consists of strengthening trace-to-trace (or trace-to-schema and schema-to-trace) links with the unitary trace assumption that learning involves two discrete stages, the first being storage of a unitary trace and the second consisting of learning to retrieve the trace reliably. First, the mathematical model will be summarized and then both the multiple and unitary trace predictions will be enumerated in the context of the task and the model's parameters.

A Mathematical Model for Two-stage Learning

It is well known that the acquisition data from most cued recall experiments conform to a three-state Markov process (e.g., Bower & Theios, 1964; Greeno, 1968; Kintsch, 1963). The mathematical model developed here is a version of the three-state Markov model used by Greeno et al. (1978) and elaborated by Brainerd, et al (1980, 1982). This two-stage learning model is presented graphically in Figure 3 and is detailed in Appendix A. Only a brief description of the model is presented here in order to facilitate the development of the specific empirical predictions outlined below.

Insert Figure 3 about here

In this model it is assumed that learning involves two discrete stages. Each stage is completed in an all-or-none manner and it is assumed that the stages occur in a fixed sequence (i.e., the first stage occurs before the second stage) and are independent (i.e., the difficulty of the second stage is unaffected by the difficulty of the first stage). On any particular trial, an item can be in one of three possible states: (I) An item can be in the "unlearned" State U in

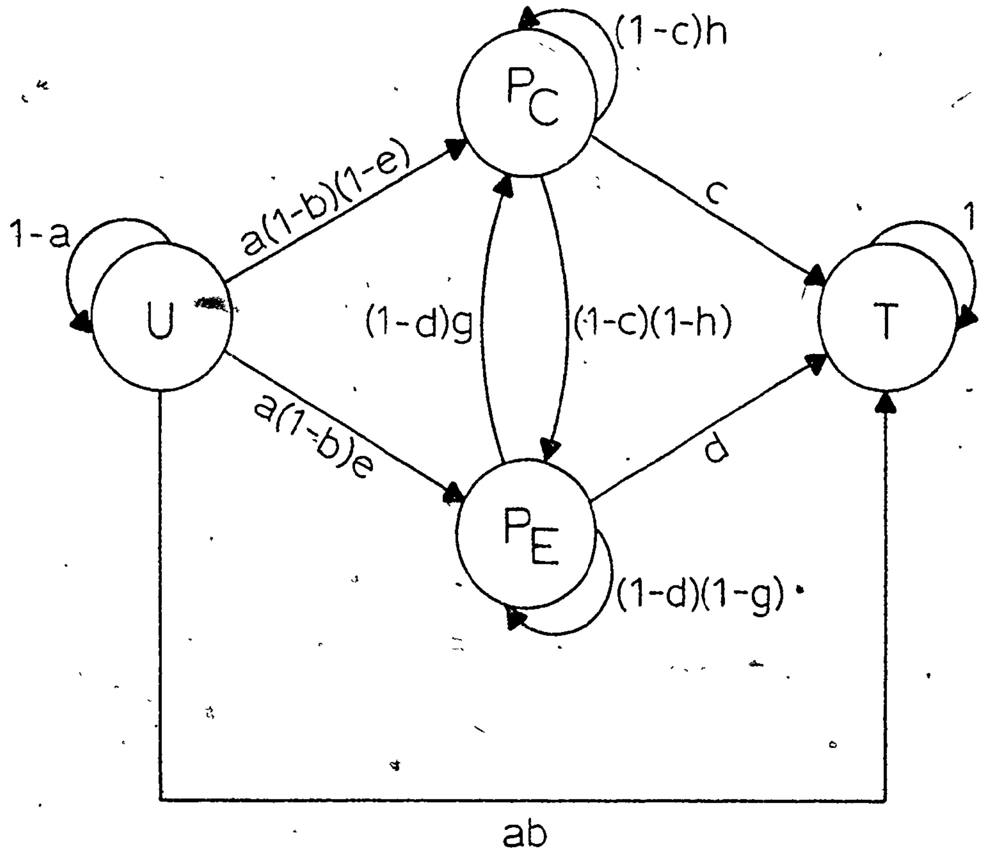

Figure 3. A Schematic Representation of the Two-Stage Markov Model of Learning.

which errors occur with probability one; (2) An item can be in the intermediate "partially learned" State P in which errors occur with an average probability 1-p and successes occur with an average probability p; or (3) An item can be in the terminal "learned" State T in which successes occur with probability one. The intermediate State P is further subdivided into two substates, P_E and P_C, depending on whether an error or success occurs, respectively.

Learning is conceptualized in terms of interstate transitions. Transitions between states have the Markov property (i.e., they occur on at most one trial, or are all-or-none) and occur only in a forward direction (i.e., there are no backward transitions as the process cannot reenter a state once it has left it). For example, subjects that start out in State U may learn by making an all-or-none transition from U to T. Alternately, they may learn by making two all-or-none transitions, one from U to P and another from P to T. Finally, subjects that start out in State P learn by making an all-or-none transition from P to T.

Associated with each state and each interstate transition are parameters that measure the difficulty of the learning process. For example, the parameters a' (i.e., 1-s) and a measure the difficulty of escaping State U on the first trial and on subsequent trials, respectively. Like all states in the process, State U is a Markov state and the probability a of leaving U is constant across trials. The value of a, therefore, gives a measure of the difficulty of whatever processes are involved in the first stage of learning. If a is large, then the first stage is easy, but if it is small then it is hard. The second stage of learning is accomplished when the item goes to State T, and may occur on the same trial as the transition from State U (i.e., a U to T transition). The parameters *b'* [i.e., *t*/(1-*s*)] and *b* are learning parameters that measure the difficulty of going directly to State T on the first learning trial and on subsequent learning trials, respectively, given the process escapes State U. The matched sets of parameters *a'*, *b'* and *a*, *b* measure the difficulty of going directly from U to T on the first trial and on subsequent trials, respectively. The value of *b* is a measure of the difficulty of the second stage of learning as it is the probability that the second stage takes no trials beyond those needed to complete the first stage. The value of 1-*b*, then, measures the probability of entering the intermediate State P on any trial after the first trial. The parameters *c* and *d* are second-stage learning parameters that measure the difficulty of escaping State P (i.e., a P to T transition) given a success or an error, respectively. Finally, *e*, *r*, *g*, and *h* are performance parameters rather than learning parameters. The larger the values of *e* and *r* the poorer performance is on the first trial in State P. The larger the values of *g* and *h* the better performance is on subsequent trials in State P.

When it is known that the data from a particular paradigm are in close agreement with this mathematical formulation, theoretical processes become more tractable, giving the investigator many strategic advantages. Previous attempts by various associative theorists to provide a mathematical characterization of cued recall have not proven successful (e.g., J. Anderson & Bower, 1973, Chapter 10; Jones, 1976, 1978a, 1978b; Ross & Bower, 1981). Indeed, few of these theorists have been interested in capturing the entire acquisition process, concerning themselves instead with predictions about mean recall early in learning (i.e., after only one or two study trials). Alternatively, as the model presented here gives an explicit structural characterization of the entire learning process, it provides the necessary quantitative framework within which precise comparisons about processes can be made. That is, not only is it possible to compare values of global performance measures (e.g., mean recall), but it is possible to evaluate hypotheses about processes involved throughout the entire acquisition sequence that contribute to this overall learning statistic.

Another strategic advantage is that this model provides a single analytical format within which the effects of different experimental manipulations can be assessed. This permits greater precision in localizing treatment effects. For example, it can be determined whether the impact of experimental maniulations occurs early in learning, late in learning, or both. In addition, it can be determined whether the variables of interest affect learning rate (i.e., speed up or slow down the rate of progress through the acquisition stages or results in stage skipping), the performance of subjects at given points during acquisition (e.g., accuracy of precriterion responding), or both (see Brainerd et al., 1982). Finally, it should be pointed out that the use of a Markov model does not entail a commitment to any particular psychological interpretation of the processes operating during learning. Rather, it provides a structural description of the processes under investigation that can be rigorously evaluated. It also serves as a basis of measurement in that the parameters of the model provide specific information about the theoretical processes under study. That is, while it has become commonplace to associate the first stage of learning with factors predominantly involved in the encoding and storage of information and the second stage of learning with factors predominantly involved in the retrieval of information (e.g., Brainerd et al., 1981; Brainerd et al., 1980; 1982; Greeno et al., 1978; Halff, 1977), this formulation does not specify the exact theoretical nature of these operations. Rather, it provides a logical framework, one that makes contact with the general information-processing view that learning processes are executed in a chronologically ordered sequence of stages (e.g., encoding → storage → retrieval), within which different theoretical formulations about these events can be tested.

Overview of the Experiments and Empirical Predictions

Three questions were addressed in the present research: (1) The nature of trace structure; that is, whether items are represented independently in separate memory traces or dependently in a single trace; (2) The nature of processes involved in acquisition; that is, whether learning consists of a gradual strengthening process or a discrete two-stage process; and (3) How preexperimental organization affects associative learning; that is, how do the parameters that measure learning and performance in early and late learning change as a function of increasing interitem similarity? First, the experiments will be outlined and then the different multiple and unitary trace predictions will be elaborated.

In the experiments reported here, a general cued recall procedure was used where the subject's task was to learn lists of word triplets. Each triad consisted of a single cue (A) and two target responses (B and C). The use of triads was motivated by the simple observation that the question of cue and target independence or dependence cannot be assessed using a standard paired-associate learning procedure as learning involves only a single associative connection. In fact, both the multiple and unitary trace models yield similar predictions when subjects learn word pairs.

In order to test the different multiple and unitary trace predictions concerning the effects of preexperimental knowledge on the operation of various learning processes, the absence, presence, and degree of interitem semantic relatedness was varied among the elements within different triads. Specifically, these effects were investigated by manipulating taxonomic category membership. Natural language categories are viewed as having a core meaning or prototype. Other members of the category vary in their degree of relationship to the prototype such that, as the judged typicality of the category member decreases, the similarity relation between that item and the prototype diminishes (Rosch, 1975). According to Rosch and Mervis (1975), differences in the processing of typical and atypical members can be explained by assuming that typical members have more features or properties in common with the prototype as well as with each other than do atypical members. In general, it has been found that (1) free recall is better for high frequency associates to a category name than for items possessing a lower rank in the norms, and (2) typicality is correlated with the ease of learning artificially constructed categories (e.g., Rosch, Simpson & Miller, 1976).

For the experiments reported here, a pool of three typical and three atypical items was selected from each of 16 categories. For each category, one typical and one atyical member was designated as the retrieval cue with the remaining two typical and two atypical items designated as repsonse terms. Cue and target typicality was manipulated in a 2 x 2 orthogonal design resulting in four list conditions: (1) typical cue and typical responses (TT); (2) tyical cue and atypical responses (TA); (3) atypical cue and typical responses (AT); and (4) atypical cue and atypical responses (AA). In the first experiment, items were drawn at random from the cue and target pools of different categories to form unrelated triplets (UR). In the second experiment, items were drawn from the cue and target pools of the same category to form related triplets that varied in the degree of interitem category membership.

Finally, in order to test the different multiple and unitary trace interpretations of the acquisition process, the experiments reported here involved a learning-to-criterion design. That is, items were considered to be learned only when subjects made two consecutive errorless passes through the entire list of triads.

The use of a criterion-learning design was important for two reasons. First, as seen earlier, the investigation of acquisition processes has typically occurred in the context of a fixed-trials design. Unfortunately, because performance in fixed-trials experiments is chiefly controlled by the difficulty of early learning stages, these investigations provide only incomplete information about the nature of the entire acquisition process, and any conclusions must be limited to these early processes. By using a stringent learning criterion in the present research, equal weight was given to both early and late learning processes. Second, while there is little doubt that, in general, preexperimental interitem organization exerts a positive effect on global measures of list acquisition (e.g., Bousfield, 1953; Cofer, 1965, 1968; Jenkins, Mink & Russell, 1958; Jenkins & Russell, 1952; Kamman, 1968; Mandler, 1967; Marshall, 1967; Postman, 1962; Postman, Frazer & Burns, 1968), the specific locus of this effect is not known. This is because all of these experiments involved fixed-trials designs which, as mentioned earlier, confound treatment effects with completeness of learning. As the goal of the present research was to determine the precise locus of these effects and, therefore, involved comparisons of lists that varied in ease of acquisition, it was particularly important to use a learning-to-criterion design because this procedure does not confound list difficulty manipulations with completeness of learning.

<u>Trace structure: Independence versus dependence.</u> The question of whether items are represented in separate memory traces or in a single trace can be evaluated by simply comparing the independent probability of acquiring each individual response with the combined probability of acquiring both responses. Given that subjects in this study must learn both the B and the C target responses associated with each stimulus cue A, current multiple trace models would depict the internal representation of each triad as consisting of either two or three independent associative links (ignoring, for the moment, the possibility of separate bidirectional links). Subjects could learn each triad in one of at least four different ways: (1) by forming two independent links between A-B and A-C; (2) by forming two independent links between A-B and B-C; (3) by forming three independent links between A-B, A-C, and B-C; or (4) by forming three independent links

connecting A, B and C to a mediating schema (as in the Ross & Bower, 1981, schema model in Figure 1). Regardless of which representation is preferred, the multiple trace interpretation makes the prediction that the probability of acquiring both targets should be a simple function of the independent probabilities of learning each response. In terms of the two-stage model, this means that the value of the parameters that measure the probability of learning both responses should be equal to the independent contribution of the parameter values associated with learning the individual responses. That is,

$$P(B \cap C) = P(B) \cdot P(C). \qquad (1)$$

The unitary trace model, on the other hand, says that learning should resemble a stochastically dependent process as members of the triad, rather than being stored in separate memory locations, are represented in a single memory trace. As the principles of relational association that operate to form pairs should be the same as those involved in learning higher-order groupings, the relationship between the parameters that measure the probability of learning both responses and learning the single responses should be one of mutual dependence as items are stored in and retrieved from the same unitary trace structure. That is,

$$P(B \cap C) = P(B) = P(C). \qquad (2)$$

<u>Acquisition processes: Strength versus all-or-none</u>. The second research question concerns the issue of the operation of processes (i.e., encoding, storage, association, and retrieval) in early and late learning. To review, in the unitary trace theory (at least in the storage-retrieval version) there are two discrete stages of learning (storage of a unitary cue-target representation and learning

to retrieve the target given the stimulus as a retrieval cue) that are executed in a fixed sequence. Associated with these stages are three states (U, P, and T) each of which correspond to a discrete level of correct response probability (zero, $\underline{p}$, and one, respectively). Transitions between states (i.e., learning) occur on at most one trial or are all-or-none, and improvements in performance occur in a discrete rather than a continuous manner.

Alternatively, in the schema version of the multiple trace theory, learning consists of (1) storing or activating the concept nodes in memory and constructing or discovering a mediating schema, and (2) of gradually incrementing the strength of trace-to-schema and schema-to-trace links.

These two interpretations lead to different expectations about how performance changes across trials. The strength interpretation leads to the prediction that changes in the probability of a correct response should increase in a continuous fashion across trials, whereas the all-or-none interpretation says that only discrete changes should be observed. Given the former interpretation, the mathematical model that was used to analyze the data should provide only a poor fit at best as this model assumes that changes in performance occur in a discrete manner. In particular, the strength assumption leads to the prediction that the probability of a correct response, $\underline{p}$, in the intermediate "partially-learned" State P should increase continuously until it reaches asymptote (i.e., unity). Conversely, the all-or-none or unitary trace interpretation leads to the prediction that, consistent with the mathematical model, changes in performance occur in a discrete manner and that the value of p should remain stationary across trials in the intermediate state.

These contrasting interpretations can be easily assessed by comparing the degree of correspondence between the predicted and observed distributions of intermediate-state statistics that are directly related to the value of p, namely, length of consecutive success runs before the last error, and length of consecutive error runs after the first success. Strong support for the unitary trace position would consist of finding a close agreement between the predicted and observed values of these distributions. Strong support for the multiple trace position would consist of finding a poor agreement between the predicted and observed values of these distributions. More importantly, the schema model makes the prediction that the degree of correspondence should vary as a direct consequence of variation in the degree of interitem relatedness. Specifically, the tendency to reject the null hypothesis of goodness-of-fit between the predicted and observed distributions should increase as list difficulty increases (i.e., TT → AT → TA → AA → UR). This is because items in more difficult conditions (i.e., unrelated or only weakly related triads start out at lower initial strength levels and, therefore, require more strenghtening throughout the course of learning than items in easier list conditions (i.e., strongly related triads).

The multiple and unitary trace interpretations also lead to two different parameter-specific predictions. According to the multiple trace theory, the probability of learning an item is said to increase after a correct response has been made. This is because the strength values associated with each of the interitem links is incremented following successful retrieval. This implies that the value of the parameter c, which measures the probability of making a transition from the intermediate state to the learned state (a P to T transition) given a correct response, should be greater than zero. According to the unitary trace model, on the other hand, the probability of learning an item is said to increase after an error has been made in the intermediate state. This is because " ... subjects rely on errors to tell them if further work needs to be done on an item ... " (Brainerd et al., 1981, p. 13). What this implies is that the value of the prameter d, which measures the probability of making a P to T transition given an error, should be greater than zero.

Loci of preexperimental organization effects. Both the multiple and unitary trace hypotheses agree that related items should be easier to learn than unrelated items. They differ, however, with respect to the locus of the effects of interitem semantic relatedness. In fact, there are four different hypotheses regarding the expected pattern of parameter changes as a function of manipulating interitem semantic relatedness. Since these hypotheses involve parameter-specific predictions, verbal definitions of the theoretical parameters of the two-stage learning model are presented in Table 1 for convenience of exposition.

Insert Table 1 about here

Recent versions of the multiple trace theory have not explicitly adopted a stage analysis of the learning process. These models do,

TABLE 1

Qualitative Definitions of the Theoretical Parameters of the Two-stage Learning Model

Stage and Parameter	Process Measured by the Parameter
First-stage learning parameters:	
	The parameters <u>a'</u> *and* <u>a</u> *measure the difficulty of the initial stage of the learning process where the probability of a correct response is zero. In psychological terms, these parameters measure the difficulty of encoding and storing the items in memory.*
<u>a'</u>	The probability of escaping the unlearned State U on the first study trial.
<u>a</u>	The probability of escaping the unlearned State U on any trial after the first study trial.
Second-stage learning parameters:	
	The parameters <u>b'</u> *and* <u>b</u> *measure the ease of second-stage learning as they give the probability that the process skips the partially-learned state and goes directly to the learned state from the unlearned state. In psychological terms, these parameters measure the difficulty of retrieval.*
<u>b'</u>	The probablity of jumping directly to the learned State T on the first study trial given that the process has escaped the unlearned State U on the first study trial (i.e., a U-to-T transition).
<u>b</u>	The probability of jumping directly to the learned State T on any trial after the first study trial given that the process has escaped the unlearned State U on any trial after the first study trial (i.e., a U-to-T transition).
	The parameters <u>c</u> *and* <u>d</u> *are also second-stage learning parameters. They give the probabilities of escaping the intermediate partially-learned State P given that the process either starts out in State P or makes a U-to-P transition. In psychological terms, these parameters measure the difficulty of retrieval learning.*
<u>c</u>	The probability of escaping the partially-learned State P following a success in State P (i.e., a P_C-to-T transition).
<u>d</u>	The probability of escaping the partially-learned State P following an error in State P (i.e., a P_E-to-T transition).

Table 1 (Continued)

Second-stage performance parameters:	
The parameters r, e, g and h are all measures of performance while the process occupies the intermediate partially-learned State P. In psychological terms, these parameters measure the difficulty of retrieval between the time an item is stored and the time that a reliable retrieval system is acquired.	
r	The probability of an error on the first trial in the intermediate partially-learned State P if the process jumps from the unlearned State U to State P on the first study trial (i.e., a U-to-P transition).
e	The probability of an error on the first trial in the intermediate partially-learned State P if the process jumps from the unlearned State U to State P on any trial after the first study trial (i.e., a U-to-P transition).
g	The probability of a success on any trial after the first trial in the intermediate partially-learned State P given an error on the preceding trial (i.e., a P_E-to-P_C transition).
h	The probability of a success on any trial after the first trial in the intermediate partially-learned State P given a success on the preceding trial (i.e., a P_C-to-P_C transition).

however, contain assumptions about the logical and temporal course of processes operating during acquisition. In fact, there are two plausible versions of this general position. In the first version (HAM), early learning consists of storing (i.e., activating and tagging) the separate concept nodes in memory that correspond to the cue and target words. This stage is relatively brief as storage is typically completed by the end of the first study trial. The difficulty of storage is affected only by whether the cues and targets are words or nonwords. Later learning consists of linking the separate cue and target concepts together by propositionalizing a relationship between the nodes. The difficulty of this stage is determined by whether or not a preexperimental relation exists between the concepts but not by the degree of interitem relatedness. According to this position, interitem relatedness should not affect the first-stage learning parameters $\underline{a}'$ and $\underline{a}$ as this stage simply consists of activating the cue and target concept nodes in memory. Rather, the effects of relatedness should be limited to the second-stage learning parameters $\underline{b}'$, $\underline{b}$, and $\underline{d}$ as it is only this later stage of learning that involves propositionalizing a relation between cue and target concepts. As the rate of learning is said to be solely contingent on the presence or absence of a preexperimental relation and not on the rank of the items in the norms, the values of these second-stage learning parameters should not change as a function of variation of the typicality of items within a triad. That is, the effects of relatedness should only be observed between unrelated (Experiment 1) and related (Experiment 2) triads.

In the second version of the multiple trace theory (schema model), early learning not only involves storing the cue and target concepts in memory, but also consists of simultaneously activating pathways emanating out from each node. If an intersection (i.e., schema) is found (or can be constructed) than the initial strength value of each link is incremented and tagged. Once a schema is found, the probability of correct responding can change in one of two ways. First, if the initial strength value associated with each of the connecting links is sufficiently high, recall will occur with probability one. In this case, learning is said to be complete. Alternately, if this strength value is relatively low, the probability of a correct response will be greater than zero but less than one. In this case, further work must be done in order to learn the item. This later learning consists of gradually incrementing the strength of each link on study trials and successful retrieval attempts, until the probability of recall reaches unity. According to the schema model, the difficulty of both early and late learning is determined not only by whether or not a preexperimental relation exists, but also by the degree of interitem relatedness. That is, the more features the cues and targets share, the greater the likelihood that a link will be located between them early in the learning process. In addition, there is a greater likelihood that the strength of this link is close to its optimum value, needing only modest increments, if any, to produce perfect recall.

The schema version of the multiple trace theory makes the prediction that relatedness effects should be observed in the learning parameters of both early and late acquisition stages (<u>a'</u>, <u>a</u>, <u>b'</u>, <u>b</u>, and <u>d</u>). This is because propositional links between concepts and the mediating schema are important from the very outset of learning. Moreover, the degree of relatedness within a triad should affect the value of these parameters as rate of learning is said to be directly influenced by the strength of preexisting relationships between concepts in memory. This assumption leads to the prediction that the value of these parameters should increase as a function of increasing interitem similarity in the following manner: (1) the typical-typical (TT) triads should be easiest as they possess the greatest overall similarity; (2) the mixed triads should be intermediate in difficulty, with the atypical-typical (AT) triplets being easier than the typical-atypical (TA) triplets because the former contains two items high in similarity whereas the latter contains two items low in similarity; (3) the atypical-atypical (AA) triads should be the hardest of the related lists as it possesses the least overall similarity between triplet members; and (4) the most difficult list of all should be the unrelated (UR) triads in Experiment 1, where preexperimental relations are entirely absent.

The unitary trace theory contains an explicit chronometric stage analysis. According to this model, early learning consists of storing a single relational representation of the cue and target in memory and learning to recognize an item as one for which a unitary trace has been stored. The difficulty of this first stage is said to be determined by two factors, (1) the ease of encoding and storing both cue and target items, and (2) the ease of developing an organized memory structure. Like the schema version of the multiple trace theory, then, the unitary trace theory predicts that the first-stage learning parameters $\underline{a}'$ and $\underline{a}$ should be affected by the degree of overall interitem similarity within the triad, as determined by the amount of feature overlap between cues and targets. Therefore, the ease of storing items in a unitary trace should vary as a function of the list difficulty ordering given above (i.e., TT → AT → TA → AA → UR).

According to the unitary trace model, later learning consists of developing a specific retrieval plan that enables the subject to locate the trace reliably upon presentation of the retrieval cue. While this second stage, like the first stage, is affected by both cue and target manipulations and degree of interitem organization, there are two possible interpretations of retrieval learning within the unitary trace framework. In the first of these, the early storage-retrieval model and the encoding-specificity model, learning to retrieve consists of fine-tuning stimulus features to match features stored in the trace. As this is essentially a cue-dependent process where the ease of learning to recall is directly related to the redintegrative power of the search cue, target manipulations should have a smaller effect than cue manipulations on this stage. This means is that the second-stage learning parameters $\underline{b}'$, $\underline{b}$, and $\underline{d}$ should be primarily affected by cue typicality but not target typicality.

According to the second version of the unitary trace theory, namely, Halff's modified storage-retrieval model, items escape the first stage of learning when subjects have acquired one of two retrieval systems. The first operation consists of a general retrieval heuristic that produces successful retrieval with a constant probability $\underline{p}$ regardless of the nature of the studied material (i.e., context-free operations). The second operation consists of a specific retrieval algorithm that produces successful retrieval with probability one and is sensitive to the nature of the studied material (i.e., context-sensitive operations). If the subject requires a specific retrieval algorithm, learning is said to be complete as recall occurs with probability one. If, on the other hand, the subject acquires a general retrieval heuristic, the process enters the second stage of learning. Here, subjects must do further work to develop a specific retrieval algorithm because heuristic retrieval operations are limited in their applicability, resulting in recall levels that are less than optimal (i.e., less than unity).

There are two predictions that can be derived from the modified storage-retrieval model, one concerning the behavior of second-stage learning parameters (<u>b'</u>, <u>b</u>, and <u>d</u>) and the other concerning the behavior of second-stage performance parameters (<u>r</u>, <u>e</u>, <u>g</u> and <u>h</u>). First, because second-stage acquisition consists of developing an item-specific algorithmic retrieval plan that involves both cue-dependent (search operations) and target-dependent (decoding operations) processes, the difficulty of accomplishing this stage should be affected by both stimulus and response manipulations. That is, because algorithmic retrieval operations are thought to be <u>sensitive</u> to the nature of the studied material, both cue typicality and target typicality should affect the rate of second-stage retrieval learning as reflected in changes in the learning parameters <u>b'</u>, <u>b</u> and <u>d</u>. Second, because recall is mediated by context-free heuristic retrieval operations while the process is in the second-stage of learning, performance in this intermediate state should not be affected by interitem semantic similarity. That is, because heuristic retrieval operations are thought to be insensitive to the nature of the studied material, the parameters *r*, *e*, *g*, and *h* that measure retrieval throughout the partially-learned state, should not vary as a function of (1) whether or not items within a triad are related (Experiment 1 versus 2), or (2) the degree of category membership (Experiment 2).

Summing up, the HAM version of the multiple trace theory makes two predictions. First, the parameters that measure the ease of first-stage learning (*a'* and *a*) should not vary as a function of typicality manipulations. That is, their values should remain invariant across experiments as well as within experiments. Second, the parameters that measure the ease of second-stage learning (*b'*, *b* and *d*) should take on greater values in conditions where a preexperimental relation exists between members of a triad (Experiment 2) than in conditions where this relation is absent (Experiment 1). In addition, the value of these second-stage learning parameters should remain invariant across conditions in which item typicality is manipulated (Experiment 2).

The schema version of the multiple trace theory, on the other hand, makes the prediction that both the presence and the degree of preexperimental interitem similarity should affect the learning parameters of both the first and second stage of acquisition (*a'*, *a*, *b'*, *b*, and *d*). In particular, the values of these parameters should increase as a function of increasing interitem similarity (i.e., UR → AA → TA → AT → TT).

Finally, both versions of the unitary trace theory predict that the values of the first-stage learning parameters (a' and a) should increase as a function of increasing interitem similarity (UR $\rightarrow$ AA $\rightarrow$ TA $\rightarrow$ AT $\rightarrow$ TT). The early storage-retrieval model and the encoding-specificity model both make the prediction that the second-stage learning parameters b', b, and d should be primarily affected by cue typicality but not target typicality. According to the modified storage-retrieval model, on the other hand, the second-stage learning parameters b', b and d should be affected by both cue and target typicality. In addition, the parameters that measure second-stage performance (r, e, g, and h) should remain invariant across manipulations of interitem similarity both between experiments and within experiments.

EXPERIMENT 1

The primary purpose of the first experiment was to test the different multiple and unitary trace predictions about item representation and the nature of processes involved in the acquisition of unrelated word triplets. In addition to providing crucial information about associative learning in the absence of preexperimental knowledge, the data from this experiment serves as a benchmark for assessing the effects of the presence of semantic relatedness on acquisition (Experiment 2).

A secondary aim of the first experiment was to determine whether or not item typicality was correlated with any factor that might influence learning other than degree of category membership. This was particularly important as the purpose of the second experiment was to obtain a pure measure of the influence of preexperimental interitem relatedness on acquisition. Despite the fact that both the typical and atypical items were equated on semantic variables that are known to affect learning (i.e., concreteness, imagery, meaningfulness, and concept familiarity), it could be that typicality was confounded with some characteristic of the words themselves that could influence performance (e.g., frequency, length, pronounceability). As category relatedness was not manipulated in this first experiment, systematic differences in learning between typical and atypical items would not be expected unless there was some extraneous variable associated with typicality.

Method

Learning lists. The triads that were used in both Experiments 1 and 2 consisted of typical and/or atypical items drawn from 16 different categories. Only basic-level category members were used as they are known to carry the most category information, possess the highest cue validity, and are the most differentiated from each other (Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976). As previous research has indicated that ratings of category typicality and category production frequency are highly correlated (Mervis, Catlin & Rosch, 1976), typicality or degree of category membership was manipulated by selecting items of high and low frequency from the Battig and Montague (1969) production frequency norms. Typical items were defined as those with a production frequency rank above 9 and atypical items were defined as those below 18. Both typical and atypical items were matched as closely as possible on concreteness (typical $\bar{x}$ = 5.90, SD = .45; atypical $\bar{x}$ = 5.82, SD = .41), imagery (typical $\bar{x}$ = 5.77, SD = .45; atypical $\bar{x}$ = 5.70, SD = .47), meaningfulness (typical $\bar{x}$ = 4.49, SD = .57; atypical $\bar{x}$ = 4.21, SD =

.44), and concept familiarity (typical $\bar{x}$ = 6.14, SD = .42; atypical $\bar{x}$ = 5.88, SD = .46) according to the Toglia and Battig (1978) semantic word norms.

A total of 48 typical and 48 atypical items were selected, 3 typical and 3 atypical members from each of the 16 different semantic categories. For each semantic category, 1 typical and 1 atypical member was designated as the retrieval cue with the 2 remaining typical and atypical items designated as the target items. For each experiment, a total of 64 triads were constructed using a 2 (cue and target) by 2 (typical and atypical) orthogonal design resulting in 4 different triad combinations for each of the 16 semantic categories. These triads were then arranged in 4 different lists, typical cue-typical targets (TT), typical cue-atypical targets (TA), atypical cue-typical targets (AT), atypical cue-atypical targets (AA), each of which consisted of 16 items, one from each of the semantic categories. In the first experiment, typical and atypical items were drawn at random from the cue and target pools of different semantic categories to form unrelated triads. In the second experiment, typical and atypical items were drawn from the cue and target pools of the same semantic category to form related triads (see Appendix D).

Subjects. The subjects were 100 introductory psychology students who participated in the experiment to fulfill a course requirement.

Design and procedure. Subjects in this experiment learned lists of unrelated triads with cue and target typicality manipulated within subjects. Each subject learned a 16 item list that consisted of 4 unique triads (randomly sampled without replacement) from each of the 4 different cue and target combinations TT, TA, AT, and AA. A total of 4 different learning lists were generated in this manner. The 100 subjects were randomly assigned to one of these four list conditions with the one stipulation that there be an approximately equal number of males and females in each condition.

The triads were presented visually using a Kodak Ektagraphic slide projector. The stimulus cue was centrally located just to the left of fixation and the two target responses were located one above the other just to the right of fixation. Learning was by the anticipation method at a 3.5 sec presentation rate. Subjects were instructed that the goal of the experiment was to learn both responses for each stimulus; that is, on each test trial they were to try to give both of the responses that were associated with each stimulus cue. In the event that they could not remember both words, subjects were instructed to try to give at least one of the two responses and were encouraged to guess. The reason that single response production was stressed in the instructions was that this procedure maximized the opportunity to observe the acquisition of each individual response, permitting a more rigorous evaluation of the question of item independence during acquisition.

In order to eliminate serial-order effects, four different list orderings were used during presentation. In order to eliminate short-term memory effects, list orderings were constrained such that at least six items intervened between test trials for a triad. The acquisition criterion was two consecutive errorless runs through the list in which both responses were given to each stimulus.

Results and Discussion

Scoring procedures. First, it should be mentioned that because a correct response could not be made on the first test trial using the anticipation method, Trial 1 in the data was always the test trial following the first study opportunity. Second, since the primary purpose of this investigation was to examine the learning of word triplets, the critical data consisted of recall protocols that were defined by the joint event of learning both the B and C responses. That is, the first scoring method consisted of strings of errors and successes where a response was considered correct if and only if the subject produced both BC responses (the combined responses BC analysis).

Finally, in order to evaluate the various multiple and unitary trace predictions, in particular the question of item independence, two additional scoring procedures were employed. Since each triad can be conceived of as containing two, more or less separate, paired associate elements (i.e., A-B and A-C), these next two scoring methods were concerned with the analysis of each individual response member. That is, the second scoring method consisted of strings of errors and successes where a response was considered correct if the subject produced the B response only (the single response B analysis), and the third scoring method consisted of strings of errors and successes where a response was considered correct if the subject produced the C response only (the single response C analysis). The B response member of the triad was arbitrarily defined as the upper target term on the presentation slide, with the C response member being defined as the lower of the two target terms.

Qualitative results. While most of the analyses rely heavily on the three-stage Markov model, some qualitative findings will be reported first. In Table 2, the descriptive statistics by list

---

Insert Table 2 about here

---

statistics by list condition for total errors during triad acquisition (combined responses BC data) are presented. A 2 (Cue: Typical versus Atypical) by 2 (Target: Typical versus Atypical) within subject analysis of variance showed that both main effects were significant [Cue: Typical vs. Atypical $F(1,24) = 5.21$, $p < .03$; Target: Typical vs. Atypical $f(1,24) = 19.41$, $p < .001$] as well as the interaction between them [$F(1,24) = 10.82$, $p < .003$]. Subsequent Newman-Keuls analysis of the interaction indicated that the list-difficulty ordering was TT < TA = AT = AA ($p < .05$). Thus, the effect of manipulating the typicality of cues and targets in the absence of category membership was negligible. The limited facilitation observed during acquisition was found to be restricted to those triads in which both the stimulus and response members were typical items.

Three-state model. The general model that was used to analyze the data is presented in Appendix A (also, see Brainerd et al., 1982). Before proceeding to specific tests of the multiple and unitary trace theories of item representation and acquisition, it must be established that the three-state Markov model provides an adequate account of both the single and combined response data. These goodness-of-fit tests will be presented first and then each specific hypothesis will be considered in turn.

Goodness of fit. Two questions must be answered in order to establish goodness of fit of the three-state model. The first

TABLE 2

Means and Standard Deviations (in brackets) for Total Errors in the Combined Response Data by List Condition in Experiment 1.

LIST CONDITION			
Typical-Typical	Typical-Atypical	Atypical-Typical	Atypical-Atypical
105.16	124.56	118.68	123.32
(20.38)	(30.37)	(27.42)	(26.38)

question asks whether the three-state model is necessary, or will a simpler (i.e., two-state) model suffice? The second question asks whether the three-state model is sufficient, or are the data more complicated than a simple three-state process?

In order to answer the first question, the three-state model must be compared to a qualitatively simpler model (Brainerd et al., 1982; Greeno, 1968; Theios, Leonard & Brelsford, 1977). Since the three-state model assumes that learning is a two-stage process, the question of necessity can be examined by comparing it to the simpler general all-or-none model that has only two states and assumes that learning is a one-stage process (Greeno & Steiner, 1964; Polson, 1970). In order to decide which model provided the best account of the data, a likelihood ratio test was performed (for details, see Brainerd et al., 1982, Equation 34; Greeno, 1968). Briefly, "the likelihood of a set of data with respect to a given theoretical model is the a posteriori probability of the data assuming that the data were in fact generated by a process isomorphic with that of the model" (Theios et al., 1977, p. 220).

To evaluate the question of necessity, the likelihood of the data was separately estimated under the assumption that learning was two stage and under the assumption that learning was one stage. Since the likelihood function for the two-stage model involved the estimation of 8 identifiable parameters, whereas the likelihood function for the one-stage model involved the estimation of 5 identifiable parameters, let L_8 symbolize the former likelihood of the data and let L_5 symbolize the latter likelihood of the data. The likelihood ratio test was computed by taking twice the negative natural log of L_5

divided by L_8.[3] This ratio tests the null hypothesis that the one-stage assumption gives as good an account of the data as the two-stage assumption. The distribution of this test statistic is approximately $\chi^2(3)$. The statistics for each response analysis by list condition appear in Table 3. The extremely large values of the $\chi^2(3)$ statistic leave little doubt that acquisition was at least a two-stage process in all list conditions for both the single and combined response analyses.

Insert Table 3 about here

As it has been established that the two-stage model is at least necessary, the second question of the sufficiency of this model can now be examined. This question concerns whether or not fine-grain statistics of the data conform to a three-state Markov process. That is, this question concerns the degree of correspondence between the observed distributions of various statistics of the data and the distributions of these same statistics as predicted by the model (see Brainerd et al., 1982; Greeno, 1968). For the two-stage model to be regarded as sufficient, it must give not only a statistically adequate account of the acquisition process as a whole, but must also provide an equally good account of the distributions of the statistics associated with each of the two learning stages. To this end, observed and predicted distributions of four statistics of the data were compared. Two of these statistics were relevant to the process as a whole (trial of last error and total errors), one statistic was relevant to the first stage of learning (errors before the first

TABLE 3

One-Stage versus Two-Stage Learning for Individual and Combined Response Data by List Condition for Experiment 1

List Condition and Response Analysis	STATISTIC $-2\log_e L_5$	$-2\log_e L_8$	$\chi^2(3)$
Typical-Typical:			
Single response B	4338.49	3898.06	440.43
Single response C	4365.21	3925.75	439.46
Combined responses BC	4265.59	3821.59	444.00
Typical-Atypical:			
Single response B	4259.24	3784.63	474.61
Single response C	4323.23	3857.34	465.89
Combined responses BC	4181.65	3706.15	475.50
Atypical-Typical:			
Single response B	4281.01	3831.47	449.54
Single response C	4195.96	3744.93	364.49
Combined responses BC	4104.41	3604.91	449.50
Atypical-Atypical:			
Single response B	4423.07	3941.13	481.94
Single response C	4439.41	3941.20	498.21
Combined responses BC	4306.12	3803.73	502.39

NOTE. The distribution of the statistic in the third column is only approximately $\chi^2(3)$. The critical value is 7.82 at $p < .05$.

success), and one statistic was relevant to the second stage of learning (total errors after the first success). [The expressions for the sampling distributions of these statistics in terms of the 8 identifiable parameters of the model were derived from matrix algorithms developed by Bernbach (1966) and Millward (1969) and can be found in Appendix B, Brainerd et al. (1982), and Greeno (1968).]

Insert Table 4 about here

For each response analysis by list condition, the predicted distribution of each statistic was obtained by substituting the maximum likelihood estimates of the relevant parameters in the appropriate sampling distribution of the statistic. The predicted and observed distributions for the data of Experiment 1 are presented in Table 4. The null hypothesis of interest is that the correspondence between the observed and predicted distributions be exact, or at least be within a statistically tolerable range. A total of 48 Kolmogorov-Smirnov tests were computed (4 lists x 3 response analyses x 4 statistics). Out of these 48 tests, only 2 null hypothesis rejections were obtained at the $\underline{p} < .05$ criterion (errors before the first success for the AT triads single response analysis B and the combined response analysis BC). When the $\underline{p} < .01$ criterion was used, the degree of correspondence between the predicted and observed distributions was statistically acceptable for all of the conditions. Given such a high degree of correspondence between the data and the model, it can be concluded that the three-state Markov model is both necessary and sufficient for both the single and combined response

TABLE 4

Predicted and Observed Distributions of Four Statistics of the Individual and Combined Response Data by List Condition for Experiment 1

List condition and response analysis	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	≥16
	Probability of k																
	Errors Before First Success																
Typical-Typical:																	
Single response B																	
Predicted	.093	.076	.138	.133	.113	.092	.073	.058	.046	.037	.029	.023	.018	.015	.012	.009	.024
Observed	.093	.078	.120	.125	.085	.085	.088	.103	.043	.038	.035	.018	.033	.013	.010	.003	.023
Single response C																	
Predicted	.098	.062	.131	.131	.113	.093	.075	.060	.048	.038	.031	.024	.020	.016	.012	.010	.026
Observed	.098	.063	.118	.113	.083	.098	.095	.100	.033	.035	.050	.018	.033	.023	.010	.010	.023
Combined responses BC																	
Predicted	.065	.058	.131	.133	.116	.096	.078	.063	.051	.041	.033	.027	.021	.017	.014	.011	.030
Observed	.065	.055	.118	.118	.075	.095	.110	.113	.039	.038	.050	.020	.038	.020	.010	.010	.030
Typical-Atypical:																	
Single response B																	
Predicted	.040	.063	.121	.122	.108	.092	.077	.064	.053	.044	.037	.031	.025	.021	.018	.015	.043
Observed	.040	.055	.140	.075	.063	.105	.093	.058	.068	.060	.068	.033	.028	.018	.020	.023	.058
Single response C																	
Predicted	.045	.072	.120	.119	.105	.090	.075	.063	.052	.044	.036	.030	.025	.021	.017	.015	.043
Observed	.045	.068	.120	.075	.065	.098	.113	.055	.078	.050	.065	.025	.033	.013	.018	.020	.063
Combined responses BC																	
Predicted	.025	.048	.114	.120	.109	.094	.079	.067	.056	.047	.039	.033	.028	.023	.019	.016	.049
Observed	.025	.038	.120	.063	.070	.103	.120	.055	.075	.065	.073	.035	.033	.018	.020	.023	.068

Table 4 (Continued)

List condition and response analysis	Probability of k																
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	≥16
Atypical-Typical:																	
Single response B																	
Predicted	.073	.061	.122	.122	.110	.093	.077	.063	.051	.042	.034	.028	.023	.019	.015	.012	.035
Observed	.073	.065	.053	.108	.128	.108	.055	.090	.100	.040	.038	.025	.028	.035	.025	.015	.015
Single response C																	
Predicted	.038	.057	.130	.133	.117	.098	.080	.065	.053	.043	.035	.028	.023	.019	.015	.012	.034
Observed	.038	.060	.085	.088	.133	.120	.073	.070	.113	.038	.030	.033	.035	.028	.015	.018	.018
Combined responses BC																	
Predicted	.020	.032	.124	.133	.119	.101	.084	.069	.057	.047	.038	.032	.026	.021	.018	.014	.042
Observed	.020	.035	.060	.083	.138	.108	.068	.090	.128	.045	.040	.038	.030	.048	.025	.018	.028
Atypical-Atypical:																	
Single response B																	
Predicted	.035	.068	.124	.125	.110	.093	.078	.064	.053	.044	.036	.030	.025	.020	.017	.014	.040
Observed	.035	.070	.070	.110	.130	.078	.080	.095	.070	.045	.043	.035	.030	.033	.030	.013	.030
Single response C																	
Predicted	.025	.046	.130	.137	.121	.102	.084	.068	.055	.045	.036	.029	.024	.019	.015	.013	.035
Observed	.025	.050	.075	.123	.145	.103	.068	.103	.063	.055	.040	.040	.030	.013	.033	.008	.025
Combined responses BC																	
Predicted	.008	.041	.122	.130	.117	.100	.084	.070	.058	.048	.039	.032	.027	.022	.018	.015	.044
Observed	.008	.035	.068	.105	.145	.078	.080	.113	.075	.058	.043	.045	.040	.033	.030	.013	.030

Errors After First Success

Typical-Typical:												
Single response B												
Predicted	.580	.195	.100	.052	.027	.014	.008	.004	.002	.001	.001	.000
Observed	.573	.195	.115	.068	.020	.013	.005	.010	.003	.000	.000	.000

Table 4 (continued)

List condition and response analysis	Probability of k: 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	≥16
Single response C																	
Predicted	.586	.182	.098	.053	.029	.016	.009	.005	.003	.002	.001	.001					
Observed	.578	.175	.120	.060	.035	.015	.008	.008	.003	.000	.000	.000					
Combined responses BC																	
Predicted	.597	.187	.098	.051	.027	.014	.008	.004	.002	.001	.001	.000					
Observed	.585	.190	.115	.055	.033	.005	.005	.008	.003	.000	.000	.000					
<u>Typical-Atypical</u>:																	
Single response B																	
Predicted	.594	.206	.100	.049	.024	.012	.006	.003	.001	.001	.000	.000					
Observed	.588	.220	.098	.050	.018	.015	.003	.003	.003	.003	.000	.000					
Single response C																	
Predicted	.582	.203	.102	.052	.026	.014	.007	.004	.002	.001	.001	.000					
Observed	.573	.213	.118	.048	.020	.015	.008	.003	.003	.003	.000	.000					
Combined responses BC																	
Predicted	.617	.190	.095	.047	.024	.012	.006	.003	.002	.001	.000	.000					
Observed	.600	.220	.095	.035	.013	.013	.005	.003	.003	.003	.000	.000					
<u>Atypical-Typical</u>:																	
Single response B																	
Predicted	.607	.175	.093	.050	.027	.015	.008	.005	.003	.001	.001	.000					
Observed	.580	.200	.108	.058	.033	.013	.008	.000	.003	.000	.000	.000					
Single response C																	
Predicted	.622	.173	.092	.050	.027	.014	.008	.004	.002	.001	.001	.000					
Observed	.603	.188	.098	.070	.028	.005	.000	.008	.003	.000	.000	.000					
Combined responses BC																	
Predicted	.654	.161	.085	.045	.024	.013	.007	.004	.002	.001	.001	.000					
Observed	.615	.188	.115	.058	.018	.008	.000	.003	.000	.000	.000	.000					

Table 4 (Continued)

List condition and response analysis	Probability of k																
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	≥16
Atypical-Atypical:																	
Single response B																	
Predicted	.589	.181	.100	.055	.031	.017	.010	.005	.003	.002	.001	.001					
Observed	.565	.208	.098	.058	.038	.025	.003	.003	.000	.000	.003	.000					
Single response C																	
Predicted	.600	.164	.096	.056	.033	.019	.011	.007	.004	.002	.001	.001					
Observed	.563	.218	.085	.048	.045	.028	.010	.003	.000	.000	.003	.000					
Combined responses BC																	
Predicted	.610	.172	.096	.054	.030	.017	.009	.005	.003	.002	.001	.001					
Observed	.580	.215	.095	.045	.038	.023	.003	.000	.000	.000	.003	.000					
Total Errors																	
Typical-Typical:																	
Single response B																	
Predicted	.050	.055	.109	.122	.116	.102	.086	.071	.058	.047	.038	.030	.024	.019	.015	.012	.032
Observed	.050	.068	.088	.123	.100	.090	.083	.078	.078	.043	.055	.030	.028	.030	.023	.008	.028
Single response C																	
Predicted	.055	.047	.101	.116	.113	.101	.087	.073	.060	.049	.040	.032	.026	.021	.016	.013	.035
Observed	.055	.055	.083	.123	.095	.080	.088	.093	.055	.048	.058	.035	.045	.023	.018	.015	.033
Combined responses BC																	
Predicted	.038	.036	.100	.119	.116	.104	.089	.075	.062	.051	.041	.033	.027	.022	.018	.014	.038
Observed	.038	.043	.080	.118	.093	.093	[illegible]	.095	.065	.053	.065	.035	.040	.028	.020	.013	.038
Typical-Atypical:																	
Single response B																	
Predicted	.020	.034	.095	.112	.109	.099	.086	.073	.062	.052	.043	.036	.030	.025	.021	.017	.052
Observed	.020	.043	.080	.098	.073	.108	.085	.073	.078	.055	.060	.040	.035	.030	.015	.025	.085

Table 4 (Continued)

List condition and response analysis	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	≥16
	Probability of k																
Single response C																	
Predicted	.020	.043	.095	.109	.106	.097	.085	.073	.061	.052	.043	.036	.030	.025	.021	.018	.052
Observed	.020	.053	.093	.080	.073	.088	.083	.100	.073	.045	.063	.045	.030	.030	.013	.025	.090
Combined responses BC																	
Predicted	.013	.022	.087	.108	.108	.099	.087	.075	.064	.054	.046	.038	.032	.027	.023	.019	.058
Observed	.013	.030	.075	.078	.063	.100	.098	.085	.080	.060	.070	.040	.045	.033	.015	.023	.095
Atypical-Typical:																	
Single response B																	
Predicted	.033	.049	.098	.112	.109	.099	.086	.073	.061	.051	.042	.035	.028	.023	.019	.016	.044
Observed	.033	.080	.050	.068	.113	.070	.093	.098	.105	.043	.063	.043	.023	.033	.028	.033	.028
Single response C																	
Predicted	.020	.037	.098	.116	.114	.103	.090	.076	.063	.052	.043	.035	.029	.023	.019	.015	.043
Observed	.020	.050	.068	.068	.115	.095	.100	.085	.108	.040	.068	.045	.028	.038	.023	.025	.025
Combined responses BC																	
Predicted	.008	.025	.091	.113	.113	.104	.091	.078	.066	.055	.046	.038	.031	.026	.021	.017	.050
Observed	.008	.040	.050	.065	.115	.088	.090	.100	.115	.045	.068	.055	.023	.043	.028	.035	.033
Atypical-Atypical:																	
Single response B																	
Predicted	.018	.041	.092	.108	.107	.099	.087	.075	.063	.053	.044	.037	.031	.025	.021	.017	.050
Observed	.018	.058	.063	.075	.098	.085	.103	.088	.085	.055	.035	.055	.038	.040	.023	.030	.050
Single response C																	
Predicted	.013	.030	.089	.111	.112	.104	.092	.079	.067	.056	.046	.038	.031	.025	.020	.017	.046
Observed	.013	.043	.073	.080	.103	.103	.070	.105	.093	.055	.050	.053	.033	.028	.030	.033	.035
Combined responses BC																	
Predicted	.005	.019	.084	.108	.110	.103	.092	.079	.068	.057	.048	.040	.033	.027	.023	.019	.055
Observed	.005	.025	.055	.073	.108	.098	.078	.103	.090	.065	.045	.060	.043	.045	.023	.033	.050

Table 4 (Continued)

List condition and response analysis	Probability of k: 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	>16
	Trial of Last Error																
Typical-Typical:																	
Single response B																	
Predicted	.050	.036	.095	.104	.094	.081	.067	.055	.044	.036	.028	.023	.018	.014	.012	.009	.024
Observed	.050	.040	.088	.083	.083	.103	.095	.070	.060	.053	.065	.033	.043	.025	.030	.008	.068
Single response C																	
Predicted	.055	.028	.090	.102	.095	.082	.068	.056	.046	.037	.030	.024	.019	.015	.012	.010	.026
Observed	.055	.033	.073	.090	.093	.098	.078	.075	.058	.050	.065	.035	.048	.035	.028	.015	.065
Combined responses BC																	
Predicted	.038	.023	.092	.106	.100	.087	.073	.060	.049	.040	.032	.026	.021	.017	.014	.011	.030
Observed	.038	.028	.073	.083	.083	.108	.090	.075	.063	.053	.070	.040	.045	.035	.030	.015	.068
Typical-Atypical:																	
Single response B																	
Predicted	.020	.024	.087	.100	.095	.084	.072	.061	.052	.043	.036	.030	.025	.021	.018	.015	.043
Observed	.020	.028	.078	.053	.080	.108	.083	.063	.085	.063	.060	.060	.053	.025	.028	.018	.105
Single response C																	
Predicted	.020	.031	.084	.096	.091	.081	.070	.060	.051	.043	.036	.030	.025	.021	.017	.015	.043
Observed	.020	.035	.090	.055	.063	.095	.090	.055	.090	.053	.063	.050	.060	.020	.025	.018	.108
Combined responses BC																	
Predicted	.013	.016	.083	.100	.097	.087	.075	.065	.055	.046	.039	.033	.027	.023	.019	.016	.049
Observed	.013	.020	.075	.048	.063	.098	.100	.058	.088	.070	.063	.058	.058	.028	.028	.018	.108
Atypical-Typical:																	
Single response B																	
Predicted	.033	.031	.088	.100	.094	.083	.071	.059	.049	.041	.034	.028	.023	.018	.015	.012	.035
Observed	.033	.045	.058	.073	.100	.073	.050	.098	.100	.055	.055	.058	.033	.050	.025	.023	.073
Single response C																	
Predicted	.020	.029	.094	.107	.101	.088	.074	.062	.051	.042	.034	.028	.023	.019	.015	.012	.034
Observed	.020	.038	.063	.063	.103	.083	.073	.088	.113	.048	.045	.055	.045	.044	.018	.028	.063

Table 4 (Continued)

List condition and response analysis	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	≥16
Combined responses BC																	
Predicted	.008	.019	.093	.109	.103	.091	.078	.066	.055	.045	.038	.031	.026	.021	.017	.014	.042
Observed	.008	.028	.053	.060	.103	.080	.058	.103	.118	.058	.050	.060	.038	.058	.028	.025	.075
Atypical-Atypical:																	
Single response B																	
Predicted	.018	.034	.086	.097	.093	.082	.071	.060	.051	.042	.035	.029	.024	.020	.017	.014	.040
Observed	.018	.048	.055	.063	.083	.065	.108	.080	.075	.055	.055	.048	.043	.058	.033	.025	.085
Single response C																	
Predicted	.013	.025	.090	.105	.100	.088	.075	.063	.052	.043	.035	.028	.023	.019	.015	.012	.034
Observed	.013	.035	.065	.075	.088	.078	.075	.088	.068	.068	.063	.050	.050	.043	.033	.025	.080
Combined responses BC																	
Predicted	.005	.017	.086	.103	.100	.089	.077	.066	.055	.046	.038	.032	.026	.022	.018	.015	.044
Observed	.005	.025	.048	.063	.085	.070	.105	.090	.075	.063	.058	.055	.048	.060	.033	.028	.085

data. More important, it can be concluded that the memorization of item pairs and item triplets was a two-stage, all-or-none process to a very close approximation.

<u>Trace structure: Independence versus dependence</u>. One of the central questions in this research is whether cues and targets are represented in independent memory traces or in a single memory trace. This question can be rephrased in terms of whether triad learning involved the independent or simultaneous acquisition of the individual B and C target items. The multiple trace theory says that the acquisition of the B response should be independent of the acquisition of the C response. This means that the probability of learning both BC responses (the combined responses BC analysis) should be equal to the probability of learning the single response B times the probability of learning the single response C (see Equation 1). The unitary trace theory, on the other hand, says that the acquisition of the B response should occur simultaneously with (i.e., be dependent on) the acquisition of the C response. This means that the probability of learning both BC responses should be the same as, or equal to, the probability of learning the single response B which in turn should be equal to the probability of learning the single response C (see Equation 2).

Because the likelihood of a given data set is simply a probability measure, the question of cue-target independence or dependence can be evaluated by investigating the relationship between the likelihoods of the individual and combined response data. Mathematically, these hypotheses can be examined by constructing two sets of likelihood ratio tests, one for the question of item independence and the other for the question of item dependence. In order to evaluate the question of item independence, the three-state likelihood of the combined response data was separately estimated for each list condition when all 8 identifiable parameters were free to vary (L_8) and when the values of the 8 parameters were fixed (L_0) according to the independence rule in Equation 1. These fixed parameter values were obtained by simply multiplying the maximum likelihood estimates of the corresponding identifiable parameters obtained from the single response B anlaysis and the single response C analysis. When these fixed parameter values are substituted in the likelihood function for the combined responses BC data, the result is a numerical estimate of the likelihood that learning both responses was generated by a stochastically independent acquisition process. The likelihood ratio test was computed by taking twice the negative natural log of L_0 divided by L_8. This ratio tests the null hypothesis that the independence restriction does not significantly reduce the likelhood of the combined response data. The distribution of this test statistic is $\chi^2(8)$. The statistic for each list condition appears in Table 5. Given the extremely large values of this statistic, there is little doubt that the null hypothesis can be rejected for each list condition. Contrary to the multiple trace theory, then, the evidence suggests that the acquisition of both responses was <u>not</u> generated by a stochastically independent process.

Insert Table 5 about here

In light of these results it would not be unreasonable to suggest

TABLE 5

Tests of Item Independence by List Condition for Experiment 1

List Condition	Statistic $-2\log_{\underline{e}} L_{\emptyset}$	$-2\log_{\underline{e}} L_{8}$	$\chi^2(8)$
Typical-Typical	4631.50	3821.59	809.91
Typical-Atypical	4419.09	3706.15	712.94
Atypical-Typical	4386.61	3604.91	781.70
Atypical-Atypical	4620.69	3803.73	816.96

NOTE. The critical value that the test statistic in the third column must exceed in order to reject the null hypothesis of item independence is 15.51 at $\underline{p} < .05$.

that the acquisition of both responses was generated by a stochastically dependent process. It is instructive, however, to investigate this hypothesis directly. Recall that the assumption that learning resembles a stochastically dependent process requires that the acquisition of the B response occur simultaneously with the acquisition of the C response. This implies that there should be three different relationships between the individual and combined response data: (1) The probability of acquiring the single response B should be equal to the probability of acquiring the single response C; (2) The probability of acquiring both BC responses should be equal to the probability of acquiring the single response B; and (3) The probability of acquiring both BC responses should be equal to the probability of acquiring the single response C.

Insert Table 6 about here

All three of these relationships can be separately tested using a single likelihood ratio procedure. The null hypothesis that any two conditions are equivalent can be evaluated by simply (1) multiplying their respective likelihoods together to form a single likelihood value L_{16}, and (2) obtaining the joint likelihood value L_8 of the same data when the conditions are combined prior to minimizing the likelihood function (see Brainerd et al., 1982, Equations 46a and 46b). The appropriate likelihood ratio test consists of taking twice the negative natural log of L_8 divided by L_{16}. The distribution of this test statistic is $\chi^2(8)$. The statistic for each hypothesis by list condition appears in Table 6. Out of a total of 12 tests, only

TABLE 6

Tests of Item Dependence by Hypothesis and List Condition for Experiment 1

Hypothesis and List Condition	Statistic		
	$-2\log_e L_8$	$-2\log_e L_{16}$	$\chi^2(8)$
P(B) = P(C):			
Typical-Typical	7825.92	7823.81	2.11
Typical-Atypical	7643.14	7641.97	1.17
Atypical-Typical	7577.74	7576.40	1.34
Atypical-Atypical	7886.16	7882.33	3.83
P(B) = P(B ∩ C):			
Typical-Typical	7726.59	7719.65	6.94
Typical-Atypical	7497.27	7490.78	6.49
Atypical-Typical	7452.24	7436.38	15.86
Atypical-Atypical	7758.63	7744.86	13.77
P(C) = P(B ∩ C):			
Typical-Typical	7751.74	7747.34	4.40
Typical-Atypical	7571.31	7563.49	7.82
Atypical-Typical	7358.71	7349.84	8.87
Atypical-Atypical	7754.22	7744.93	9.29

NOTE. The critical value that the test statistic in the third column must exceed in order to reject the null hypothesis of item dependence is 15.51 at $p < .05$.

one significant rejection was obtained at $p < .05$ [$P(B) = P(B \cap C)$ for the AT list] and not one was obtained at $p < .01$. Given such a low rejection rate it is not unreasonable to conclude that the acquisition of both responses was generated by a stochastically dependent process. This result, along with the rejection of the hypothesis of item independence, provides strong support for the unitary trace hypothesis that items in a list learning situation are stored in a single memory structure.

Acquisition processes: Strength versus all-or-none. A second question in this research concerns whether learning consisted of a gradual strengthening process or consisted of two discrete changes in performance. There are two lines of evidence that are important in the evaluation of this question, namely, the overall goodness of fit of the three-state model to the data, and, more particularly, the goodness of fit of the model to intermediate-state statistics.

With regard to the question of overall goodness of fit, the schema version of the multiple trace theory predicts that the three-state Markov model should fail to provide an adequate account of the structure of the acquisition data. This is because the schema model assumes that learning consists of a single continuous strengthening process, a system which turns out to be more complicated than a simple two-stage all-or-none process. The unitary trace theory, on the other hand, assumes that learning is a three-state process in which interstate transitions occur in a discrete manner. According to this assumption, the Markov model should provide an accurate description of the data.

As reported earlier, the two-stage Markov model provided both a necessary and a sufficient account of the data. That is, it has been shown that there are exactly three performance states (U, P, and T) each of which correspond to a discrete level of correct responding (zero, $\emptyset < \underline{p} < 1$, and one), and that transitions between states (i.e., learning) are all-or-none. In general, such a finding is inconsistent with learning models "that imply that the change in the probability of a correct response across trials is either a simpler or a more complex process than two all-or-none stages" (Brainerd et al., 1982, p. 638). More specifically, this finding is inconsistent with the strength assumption in the schema version of the multiple trace model.

The results of the overall goodness-of-fit analyses clearly support the unitary trace view of how acquisition proceeds. A strength interpretation of these data might be salvaged given the post-hoc assumption that some kind of threshold process was operating throughout learning. For example, even though interstate transitions were found to be discrete, it could be argued that the actual response strength leading up to this all-or-none change in performance varied in a continuous manner until some critical response threshold was attained. Specifically, the discrete changes in correct response probability (zero to $\underline{p}$, zero to one, and $\underline{p}$ to one) that were observed during learning when the process made state-to-state transitions (U to P, U to T, and P to T, respectively) may have in fact been mediated by a continuous buildup of response strength.

What this hypothesis implies is that the observed probablity of a correct response, rather than remaining constant across trials within a state, should vary continuously until a discrete-change threshold is reached. Since, mathematically, the two-stage model assumes that performance remains stationary across trials in each state, this question can be answered by examining fine-grained aspects of the data. According to the strength assumption, the probability of a correct response should be nonstationary (i.e., it should increase in a continuous fashion) and the model should fail to accurately predict the observed values of fine-grained statistics of the data. Conversely, the all-or-none or unitary trace assumption leads to the prediction that the probability of a correct response should remain stationary within states and the model should accurately predict these data.

It has already been shown that the distributions of four statistics of the data conform to the predictions of the two-stage model. While such evidence is consistent with the unitary trace position, it could be argued that only one of these distributions (total errors after first success) provides an adequate test of the strength hypothesis. This is because three out of the four distributions involved data from a state in which the probability of a correct response was fixed. That is, because the probability of a correct response is by definition zero in the initial "unlearned" State U and by definition one in the terminal "learned" State T, continuous variation in response strength cannot be observed in these states. However, such variation is permitted in the intermediate "partially-learned" State P as the probability of a correct response p can assume values ranging between zero and one. The question of stationarity can best be answered, therefore, by examining the distribution of various statistics of the data in State P.

Insert Figures 4-7 and Table 7 about here

In light of the above arguments, the distribution of two additional State P statistics were investigated. Both of these statistics, namely, length of consecutive success runs before the last error and length of consecutive error runs after the first success, depend directly on the mathematical assumption that p remains stationary across learning trials in the intermediate state. The predicted and observed values of these distributions are shown in Figures 4-7 for the combined responses BC analyses and in Table 7 for the single response B and single response C analyses. Out of a total of 24 Kolmogorov-Smirnov tests (4 lists x 3 response analyses x 2 statistics) not one null hypothesis rejection was obtained ($p < .05$). That is, the observed distributions of these statistics were accurately predicted from the two-stage model. These data are inconsistent with any model that posits that response probability varies continuously rather than discretely across trials. More importantly, this evidence contravenes the strength assumption in the schema version of the multiple trace theory, including the post-hoc suggestion that some kind of threshold process was operating throughout learning.

Loci of preexperimental organization effects: The manipulation of typicality in the absence of category membership. A third question in this research concerns the locus of the effects of preexperimental interitem semantic relatedness on learning. The issue in the first experiment was whether or not item typicality was correlated with any

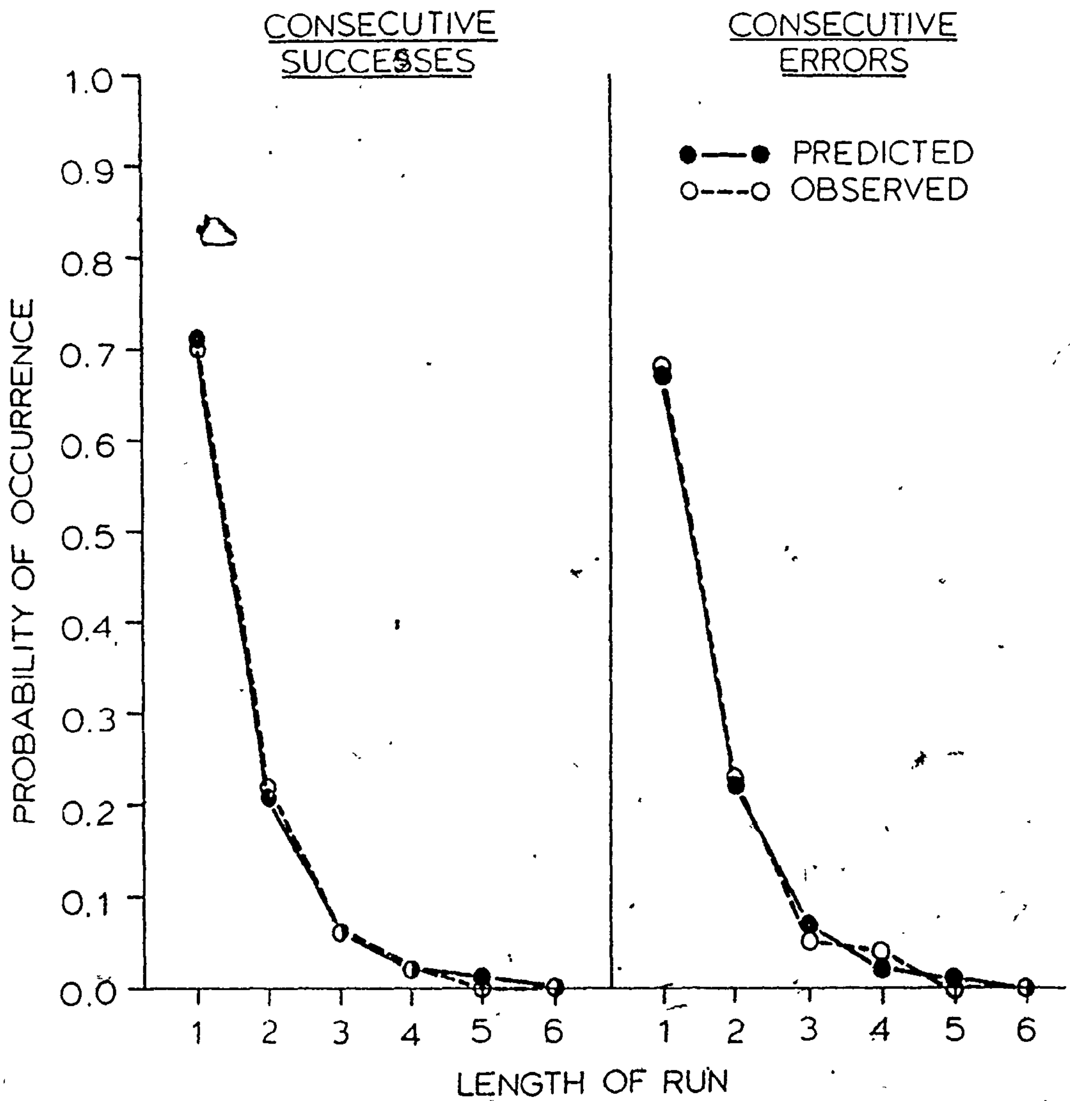

Figure 4. Predicted and Observed Distributions of Two Intermediate State Statistics, Length of Success Runs Before the Last Error (Left Panel) and Length of Error Runs After the First Success (Right Panel), for the Combined Response Data from the Typical-Typical Triad Condition in Experiment 1.

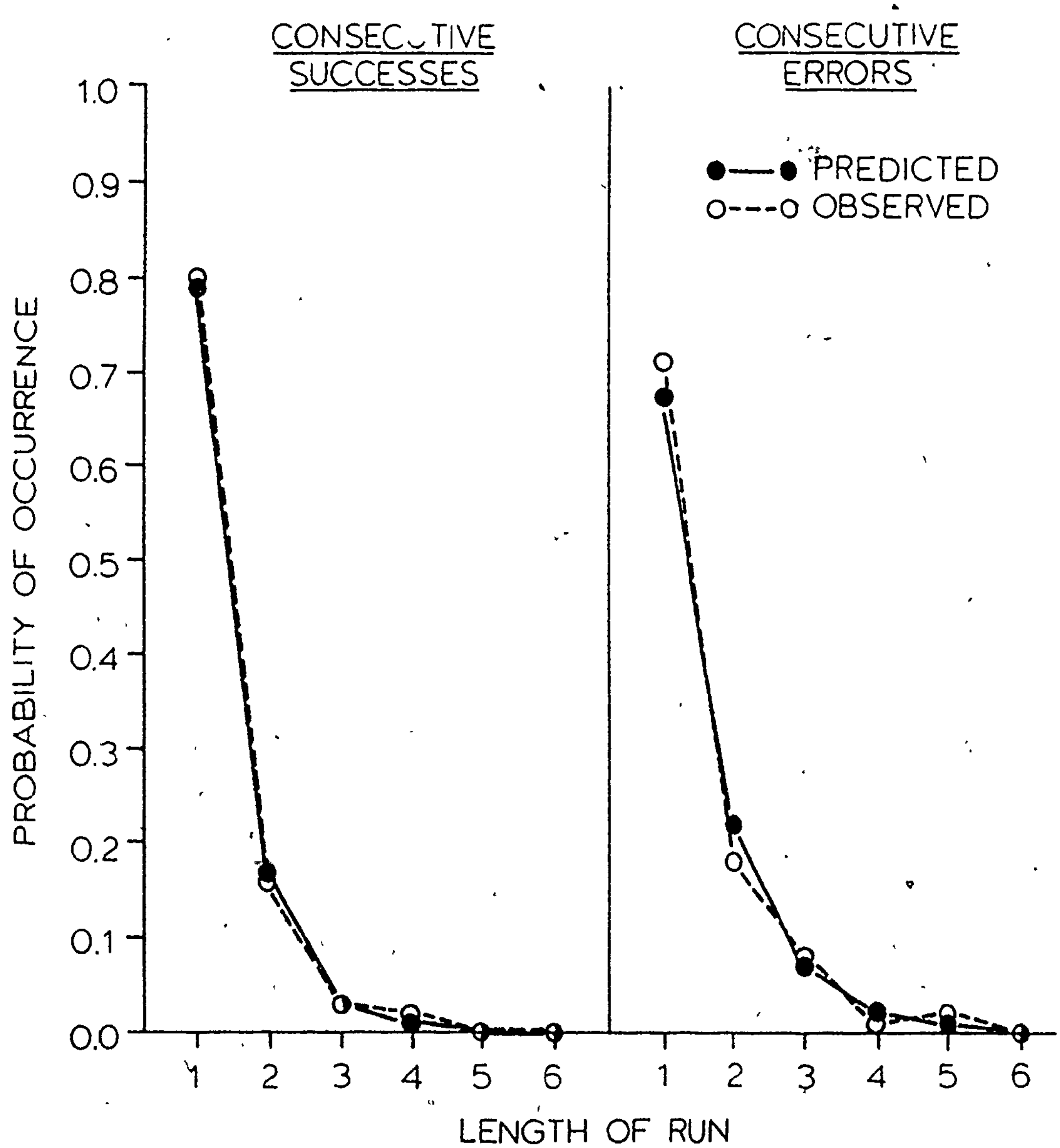

Figure 5. Predicted and Observed Distributions of Two Intermediate State Statistics, Length of Success Runs Before the Last Error (Left Panel) and Length of Error Runs After the First Success (Right Panel), for the Combined Response Data from the Typical-Atypical Triad Condition in Experiment 1.

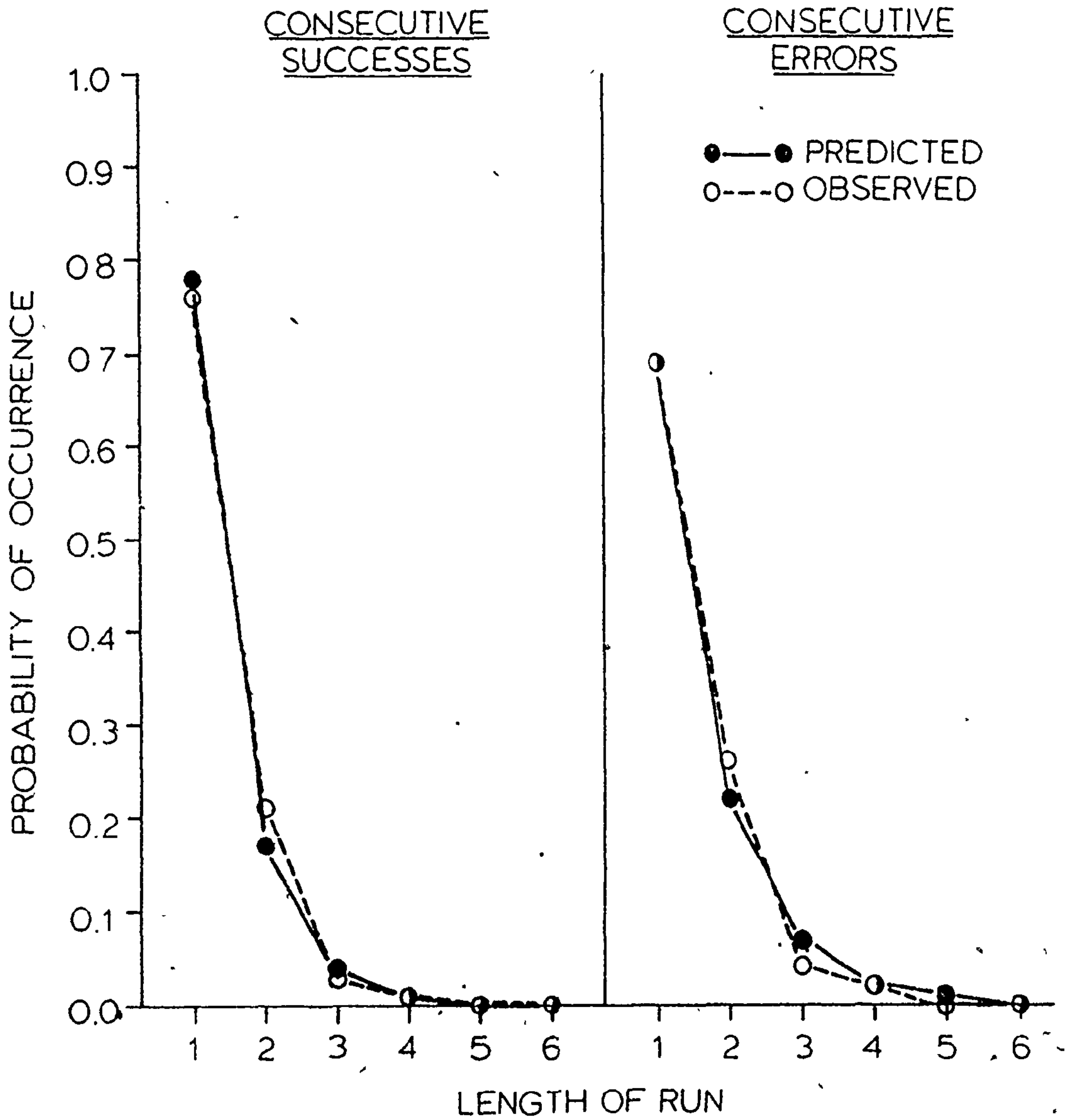

Figure 6. Predicted and Observed Distributions of Two Intermediate State Statistics, Length of Success Runs Before the Last Error (Left Panel) and Length of Error Runs After the First Success (Right Panel), for the Combined Response Data from the Atypical-Typical Triad Condition in Experiment 1.

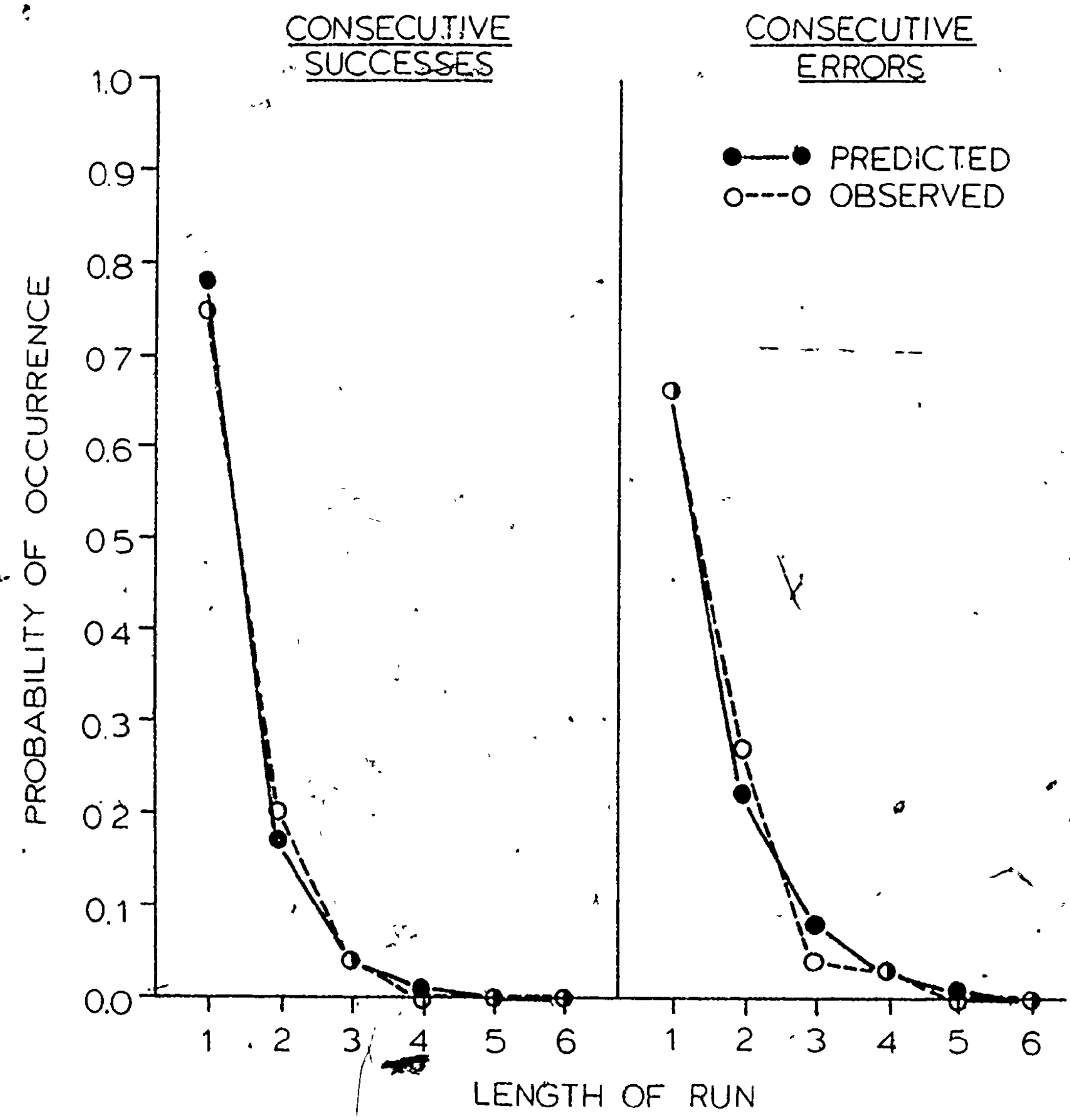

Figure 7. Predicted and Observed Distributions of Two Intermediate State Statistics, Length of Success Runs Before the Last Error (Left Panel) and Length of Error Runs After the First Success (Right Panel), for the Combined Response Data from the Atypical-Atypical Triad Condition in Experiment 1.

TABLE 7

Predicted and Observed Distributions of Two Intermediate-State Statistics of the Individual Response Data by List Condition for Experiment 1.

List condition and response analysis	Probability of k 1	2	3	4	5	6	7	≥8
	Length of Consecutive Success Runs Before the Last Error							
Typical-Typical:								
Single response B								
Predicted	.676	.219	.071	.023	.008	.002	.001	.000
Observed	.664	.225	.079	.032	.000	.000	.000	.000
Single response C								
Predicted	.699	.211	.063	.019	.006	.002	.001	.000
Observed	.692	.210	.075	.024	.000	.000	.000	.000
Typical-Atypical:								
Single response B								
Predicted	.755	.185	.045	.011	.003	.001	.000	.000
Observed	.757	.185	.041	.014	.005	.000	.000	.000
Single response C								
Predicted	.763	.181	.043	.010	.002	.001	.000	.000
Observed	.764	.180	.043	.013	.000	.000	.000	.000
Atypical-Typical:								
Single response B								
Predicted	.736	.195	.051	.014	.004	.001	.000	.000
Observed	.727	.217	.043	.013	.000	.000	.000	.000
Single response C								
Predicted	.741	.192	.050	.013	.003	.001	.000	.000
Observed	.713	.229	.054	.005	.000	.000	.000	.000
Atypical-Atypical:								
Single response B								
Predicted	.746	.190	.048	.012	.003	.001	.000	.000
Observed	.722	.218	.048	.012	.000	.000	.000	.000
Single response C								
Predicted	.739	.193	.051	.013	.003	.001	.000	.000
Observed	.720	.218	.055	.004	.004	.000	.000	.000

2 OF/DE 2

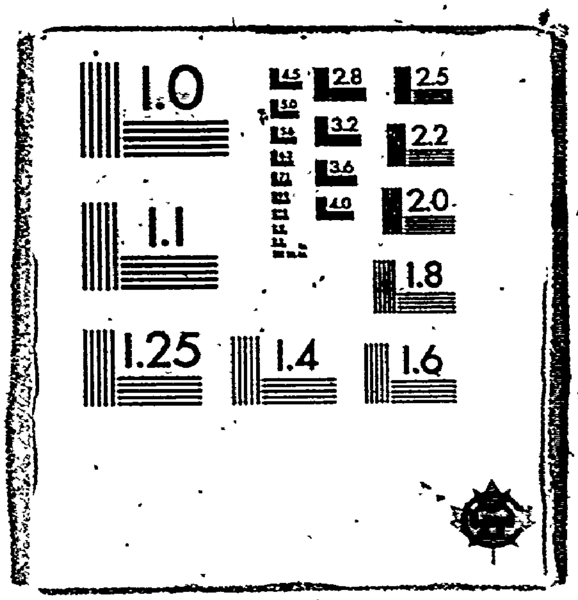

Table 7 (Continued)

List condition and response analysis	1	2	3	4	5	6	7	≥ 8
	Length of Consecutive Error Runs After the First Success							
Typical-Typical:								
Single response B								
Predicted	.700	.210	.063	.019	.006	.002	.001	.000
Observed	.696	.225	.051	.028	.000	.000	.000	.000
Single response C								
Predicted	.672	.221	.072	.024	.008	.003	.001	.000
Observed	.672	.225	.059	.040	.004	.000	.000	.000
Typical-Atypical:								
Single response B								
Predicted	.694	.213	.065	.020	.006	.002	.001	.000
Observed	.730	.171	.072	.009	.014	.000	.005	.000
Single response C								
Predicted	.677	.219	.071	.023	.007	.002	.001	.000
Observed	.713	.189	.064	.004	.026	.000	.004	.000
Atypical-Typical:								
Single response B								
Predicted	.659	.225	.077	.026	.009	.003	.001	.000
Observed	.671	.247	.048	.017	.013	.004	.000	.000
Single response C								
Predicted	.673	.220	.072	.024	.008	.003	.001	.000
Observed	.682	.223	.058	.022	.000	.005	.000	.000
Atypical-Atypical:								
Single response B								
Predicted	.669	.221	.073	.024	.008	.003	.001	.000
Observed	.673	.250	.040	.024	.008	.004	.000	.000
Single response C								
Predicted	.659	.225	.077	.026	.009	.003	.001	.000
Observed	.658	.253	.055	.027	.004	.004	.000	.000

factor that might influence learning other than degree of category membership. Because typicality was manipulated in the absence of category membership in this experiment, this question can be easily answered by comparing the relative difficulty of triplet-learning between the various list conditions.

Prior to conducting planned or unplanned between-condition comparisons, it is necessary to obtain global statistical evidence that parameter values differ between list conditions for the experiment as a whole. This test, called an "experimentwise" analysis (Brainerd et al., 1980), is an analog of the omnibus F test and is similar to the likelihood ratio procedure that was used earlier to evaluate the hypothesis of stochastic dependence. The null hypothesis that there were no differences in learning as a function of manipulating typicality in the absence of category membership was tested by simply: (1) multiplying the separate combined responses BC likelihoods of each list condition together to form a single likelihood value L_{32}; and (2) obtaining the joint likelihood value L_8 of the same data by pooling the data from all of the list conditions together prior to minimizing the likelihood function (see Brainerd et al., 1982, Equations 45a and 45b). The appropriate likelihood ratio test consists of taking twice the negative natural log of L_8 divided by L_{32}. The distribution of this test statistic is $\chi^2(24)$ and the numerical result was $\chi^2(24) = 33.64$, $p > .05$. This result indicates that there were no differences in learning between the different list conditions. More importantly, it shows that by itself, item typicality has little or no effect on learning when category membership is absent.

<u>Identification and estimation of the theoretical parameters of the two-stage model.</u> Despite a failure to find differences between list conditions, there are still several interesting questions that can be asked about the theoretical processes involved in learning word triplets. First, it is important to determine the numerical estimates of the theoretical parameters that measure learning and performance in the different acquisition stages. Second, once these parameter values have been defined, the question posed earlier of whether subjects are more likely to learn following a success or following an error can be examined. Finally, a series of within-condition parameter-invariance tests can be conducted in order to determine the relative ease with which subjects were able to store and retrieve word triplets.

In order to test these parameter-specific hypotheses about theoretical processes it is necessary to resolve a technical problem, namely, locating an identifying restriction. Briefly, the problem is that the theoretical model (shown in Figure 3) assumes that there are 10 parameters (i.e., 10 independent dimensions on which the target data vary), whereas mathematically it has been shown that there are actually only 8 parameters that can be estimated from the data (see the identifiable model in Appendix A, and discussions in Brainerd et al., 1982; Greeno, 1968). This problem can be resolved by finding an identifying restriction on the 10 theoretical parameters that "(a) reduces the freedom of these parameters to vary (i.e., which stipulates that the <u>actual</u> dimensions of variation in the data are fewer than 10) and (b) whose validity can be evaluated with certain statistical tests" (Brainerd & Howe, 1982, p. 17).

In order to determine the testability of a restriction on the theoretical model, the restriction must be inserted into the system of equations that maps the 10 theoretical parameters with the 8 identifiable parameters (see Appendix A; Brainerd et al., 1982, Equations 25-30). If this theoretical restriction implies a corresponding restriction on the identifiable parameters, then it is testable. Because the 8 parameters of the identifiable model are estimated by the method of maximum likelihood, any testable identifying restriction can be evaluated using likelihood ratio procedures (see Brainerd et al., 1982, Equations 42a-43b).

The identifying restriction that was tested for all of the list conditions in Experiment 1 was: $\underline{r} = \underline{c} = \emptyset$, and $1-\underline{a} \geq \underline{a}'\underline{b}'(1-\underline{a}'\underline{b}')$. In terms of the identifiable model, this restriction implies that $\pi \leq \theta$. The validity of this restriction was evaluated by obtaining the likelihood of the data when all 8 identifiable parameters were free to vary, L_8, and by obtaining the likelihood of the same data subject to the identifying restriction, L_7. The appropriate likelihood ratio test of the null hypothesis that this restriction did not significantly reduce the likelihood of the data consisted of taking twice the negative natural log of L_7 divided by L_8. The distribution of this test statistic is $\chi^2(1)$ and the numerical result for all list conditions was $\chi^2(1) = 0.00$, $\underline{p} > .05$. This result indicates that this restriction was valid for all 4 list conditions in Experiment 1. The numerical values of the storage parameters, retrieval-learning parameters, and retrieval-performance parameters for the combined responses BC data under the identifying restriction appear in the first 4 rows of Table 8. The definitions of these theoretical parameters in terms of the identifiable parameters given this restriction are in Appendix C. The only important point about this restriction is that it reduces the number of retrieval-learning parameters from four to three (namely, b', b, and d) and it reduces the number of retrieval-performance parameters from four to three (namely, e, g, and h).

Insert Table 8 about here

Now that the values of the theoretical parameters have been defined, the question of whether subjects are more likely to learn following a success or following an error can be examined. Recall that in the schema version of the multiple trace model the probability of learning an item is said to increase following a correct response. That is because the strength values associated with interitem links are incremented following successful retrieval. This implies that the value of the parameter c, which measures the probability of making a transition from the intermediate state to the learned state given a correct response, should be greater than zero. As it turned out, this hypothesis was tested in conjunction with the evaluation of the validity of the identifying restriction. Because each component of the identifying restriction must be true in order for the restriction to be valid, it follows that c could not have been greater than zero for these data. Contrary to the multiple trace prediction, then, the probability of learning an item did not increase as a function of successful retrieval attempts.

The unitary trace model, on the other hand, proposed that the probability of learning an item should increase following an error.

TABLE 8

Estimates of the Theoretical Parameters of the Combined Response Data by List Condition for Experiments 1 and 2.

Experiment and list condition	First-Stage Learning Parameters		Second-Stage Learning Parameters			Second-Stage Performance Parameters		
	a'	*a*	*b'*	*b*	*d*	*g*	*h*	*1-e*
Experiment 1:								
Typical-Typical	.07	.20	.58	.13	.47	.39	.29	.22
Typical-Atypical	.03	.16	.50	.10	.50	.34	.21	.24
Atypical-Typical	.02	.18	.38	.11	.47	.42	.22	.09
Atypical-Atypical	.01	.18	.67	.10	.44	.40	.22	.15
Experiment 2:								
Typical-Typical	.40	.49	.73	.62	.66	.43	.29	.45
Typical-Atypical	.15	.37	.90	.60	.62	.33	.25	.66
Atypical-Typical	.25	.47	.84	.54	.71	.36	.27	.25
Atypical-Atypical	.09	.41	.81	.36	.51	.29	.21	.05

NOTE. These estimates were obtained under the identifying restriction that $r = c = \emptyset$ and $1 - a' \geq a'b'(1-a'b')$.

This is because an incorrect response tells the subject that further work must be done in order to learn that item. This hypothesis implies that the value of the parameter d, which measures the probability of making a transition from the intermediate state to the learned state following an error, should be greater than zero. As can be seen from Table 8 the value of d is substantially greater than zero for all list conditions. When tested statistically using standard likelihood ratio procedures, the null hypothesis that $d = 0$ was rejected in every case, $\chi^2(1)$ 10.00, $p < .05$. Consistent with the unitary trace model, then, the probability of learning increased following an error in the intermediate state.

In previous research (Brainerd & Howe, 1982; Brainerd et al., 1980; 1981, 1982), certain ordinal relationships have been observed among the various parameters within list conditions regardless of the nature of the material being memorized. Specifically, two theoretically important list-invariant relationships have been reported: (1) $\{a', a\} \leq \{b', b, d\}$; and (2) $g \leq h$. With regard to (1), recall that a' and a are measures of the difficulty of first-stage learning, whereas b', b, and d are measures of the difficulty of second-stage learning. Likelihood ratio tests revealed that $\{a', a\} <$ $\{b', d\}$ for all four list conditions in Experiment 1. These results suggest that, in general, first-stage learning was more difficult than second-stage learning. These findings are consistent with previous research in the area of paired-associate learning where storage of information was found to be more difficult than learning to retrieve that information. More importantly, these data are inconsistent with the assumption in the HAM version of the multiple trace theory that information storage plays a relatively minor role in the acquisition process.

With regard to (2), recall that g and h are second-stage performance parameters that measure the probability of a correct response following an error and following a success, respectively. The usual finding has been that the probability of a success is greater when it is preceded by a success than when it is preceded by an error. This result has been interpreted in terms of a priming effect in the intermediate state (e.g., Brainerd et al., 1980). This finding is also consistent with, although not a necessary concommitant of, the assumption in the schema version of the multiple trace theory that associations are strengthened following successful retrieval. Contrary to previous research, however, likelihood ratio tests revealed the opposite relationship, namely, $g > h$, for all four list conditions in Experiment 1. The finding, that there was a higher probability of a correct response following an error than following a success, is inconsistent with the multiple trace claim. Rather, these data are more compatible with the unitary trace suggestion that errors cue subjects to do further work in order to learn the items.

EXPERIMENT 2

The results of the first experiment provide support for a unitary trace conception of learning unrelated associative clusters. There was clear evidence that (1) members of each triad were represented dependently in a single memory trace, (2) performance remained stationary across trials within each acquisition stage, and (3) learning consisted of two discrete all-or-none changes in response probability. In addition, it was shown that learning was not systematically affected by variation in item typicality in the absence of category membership.

The purpose of the second experiment was to extend the generality of the above findings and to test the different multiple and unitary trace predictions about the acquisition of related word triplets. The issues of item representation and the nature of the processes involved in list acquisition were again foremost. In addition, the effects on acquisition of variation in the degree of preexperimental semantic relatedness were investigated. These effects were examined at both a global and a stage-specific level. At the global level, interest centers on the relationship between degree of category membership and the overall list-difficulty ordering. At the stage-specific level, interest centers on the relationship between changes in the values of different learning and performance parameters and the manipulation of cue versus target typicality.

Method

Subjects. The subjects were 100 introductory psychology students who participated in the experiment to fulfill a course requirement.

Design and procedure. The design and procedure of the second experiment was the same as that used in Experiment 1 except that subjects learned lists of related triads. Typicality was manipulated within subjects where each subject learned a 16-item list that consisted of 4 triads from each cue-target combiantion TT, TA, AT, and AA. Triads were presented visually using a Kodak Ektagraphic slide projector. The cue was centrally located just to the left of fixation and the two targets were located one above the other just to the right of fixation.

Learning was by the anticipation method at a 3.5 sec presentation rate. Four different list orderings were used during presentation and at least six items intervened between test trials for a triad. The acquisition criterion was two consecutive errorless passes through the list in which both responses were given to each stimulus.

Results and Discussion

Scoring procedures. The same scoring methods were used here as were used in Experiment 1. Trial 1 in the data was always the test trial following the first study opportunity. Recall protocols were analyzed according to three scoring criteria (1) combined responses BC analysis, (2) single response B analysis, and (3) single response C analysis.

Insert Table 9 about here

Qualititative results. The descriptive statistics by list condition for total errors during triad acquisition (combined responses BC data) are shown in the second row of Table 9. A 2 (Cue: Typical versus Atypical) by 2 (Target: Typical versus Atypical) within subject analysis of variance showed that both main effects were significant [Cue: Typical vs. Atypical $F(1,24) = 26.61$, $p < .00003$; Target: Typical vs. Atypical $F(1,24) = 125.75$, $p < .000001$] but that the interaction was not. These results indicate that typicality, when manipulated in the presence of category membership, facilitated acquisition such that typical category members were easier to learn than atypical category members. Moreover, manipulating the degree of category membership on the cue versus the target components of the

TABLE 9

Means and Standard Deviations (in brackets) for Total Errors in the Combined Response Data by List Condition in Experiment 2.

LIST CONDITION			
Typical-Typical	Typical-Atypical	Atypical-Typical	Atypical-Atypical
28.04	45.92	34.28	54.84
(8.67)	(12.49)	(9.44)	(15.32)

triads had an equivalent and additive effect on learning.

Three-state model. Before proceeding to tests of specific hypotheses concerning the acquisition of semantically related associative clusters, the goodness of fit of the three-state Markov model must be established for these data.

Goodness of fit. First, the necessity of the model was assessed. The likelihood ratio tests for the one-stage versus the two-stage learning models appear in Table 10. Because each test produced a null hypothesis rejection, it was concluded that the three-state model was at least necessary for both the single and combined response data.

Insert Table 10 about here

Second, the sufficiency of the model was assessed. The predicted and observed distributions of the same four statistics used in Experiment 1 were examined and are shown in Table 11. Of the 48 Kolmogorov-Smirnov tests that were computed (4 lists x 3 response analyses x 4 statistics), only one null hypothesis rejection was obtained at the $p < .05$ criterion (trial of last error for the AT triads single response analysis B). When the $p < .01$ criterion was used, the degree of correspondence between the predicted and observed distributions was statistically acceptable for all of the conditions. As in the first experiment, the degree of correspondence between the data and the model leads to the conclusion that the three-state Markov model is both necessary and sufficient for both the single and combined response data. Moreover, it can be concluded that the memorization of both unrelated and related associative clusters was a

TABLE 10

One-Stage versus Two-Stage Learning for Individual and Combined Response Data by List Condition for Experiment 2.

List condition and response analysis	Statistic $-2\log_e L_5$	$-2\log_e L_8$	$\chi^2(3)$
Typical-Typical:			
Single response B	2238.24	2107.46	130.78
Single response C	2223.65	2096.75	126.90
Combined responses BC	2290.55	2173.68	116.87
Typical-Atypical:			
Single response B	2632.41	2476.92	155.49
Single response C	2649.95	2460.95	189.00
Combined responses BC	2677.37	2512.54	164.83
Atypical-Typical:			
Single response B	2194.25	2021.15	173.10
Single response C	2195.25	2041.47	153.78
Combined responses BC	2215.38	2063.94	151.44
Atypical-Atypical:			
Single response B	2631.68	2495.54	136.14
Single response C	2619.57	2437.79	181.78
Combined responses BC	2566.26	2407.87	158.39

NOTE. The distribution of the statistic in the third column is only approximately $\chi^2(3)$. The critical value is 7.82 at $p < .05$.

two-stage, all-or-none process to a very close approximation.

Insert Table 11 about here

<u>Trace structure: Independence versus dependence</u>. As in the first experiment, the question of whether cues and targets are represented in independent memory traces or in a single memory trace, was assessed by examining whether triad learning involved the independent or simultaneous acquisition of the individual B and C targets. The likelihood ratio tests for item independence are shown in Table 12. Given the extremely large values of the $\chi^2(8)$ statistic, there is little doubt that the null hypothesis of item independence can be rejected for each list condition. Contrary to the multiple trace position, the evidence once again suggests that the acquisition of both responses was <u>not</u> generated by a stochastically independent process.

Insert Table 12 about here

The likelihood ratio tests for item dependence are shown in Table 13. Of the 12 $\chi^2(8)$ tests (3 component hypotheses x 4 list conditions), none produced a null hypothesis rejection at or beyond the $p < .05$ level. This finding leaves little doubt that the acquisition of both responses was generated by a stochastically dependent process. This result, coupled with the rejection of the hypothesis of item independence, once again provides strong support for the unitary trace hypothesis. It is not unreasonable to conclude,

TABLE 11

Predicted and Observed Distributions of Four Statistics of the Individual and Combined Response Data by List Condition for Experiment 2.

List condition and response analysis	Probability of k: 0	1	2	3	4	5	6	7	8	9	10	11	≥12
	Errors Before First Success												
Typical-Typical:													
Single response B													
Predicted	.445	.211	.154	.089	.048	.025	.013	.007	.004	.002	.001	.001	.000
Observed	.445	.203	.173	.058	.063	.033	.013	.010	.003	.000	.000	.000	.000
Single response C													
Predicted	.453	.218	.153	.085	.045	.023	.012	.006	.003	.002	.001	.000	.000
Observed	.453	.210	.178	.055	.050	.033	.013	.008	.003	.000	.000	.000	.000
Combined responses BC													
Predicted	.398	.232	.169	.096	.051	.027	.014	.007	.004	.002	.001	.001	.000
Observed	.398	.225	.185	.065	.065	.033	.015	.010	.003	.000	.000	.000	.000
Typical-Atypical:													
Single response B													
Predicted	.200	.265	.192	.127	.081	.051	.032	.020	.012	.008	.005	.003	.005
Observed	.200	.233	.228	.133	.080	.048	.023	.008	.030	.008	.005	.000	.005
Single response C													
Predicted	.203	.311	.179	.110	.070	.045	.029	.019	.012	.008	.005	.003	.005
Observed	.203	.275	.203	.123	.083	.040	.025	.010	.023	.010	.003	.000	.005
Combined responses BC													
Predicted	.145	.274	.205	.137	.088	.056	.035	.022	.014	.009	.006	.004	.005
Observed	.145	.238	.243	.150	.078	.060	.025	.015	.030	.008	.005	.000	.005

TABLE 11 (Continued)

List condition and response analysis	Probability of k: 0	1	2	3	4	5	6	7	8	9	10	11	≥12
Atypical-Typical:													
Single response B													
Predicted	.315	.216	.201	.122	.068	.037	.020	.010	.006	.003	.002	.001	.000
Observed	.315	.220	.155	.173	.070	.023	.023	.018	.005	.000	.000	.000	.000
Single response C													
Predicted	.310	.240	.203	.117	.062	.033	.017	.009	.005	.002	.001	.001	.000
Observed	.310	.245	.170	.140	.068	.035	.020	.010	.003	.000	.000	.000	.000
Combined responses BC													
Predicted	.248	.231	.222	.136	.076	.041	.022	.012	.006	.003	.002	.001	.001
Observed	.248	.235	.180	.178	.083	.030	.023	.018	.008	.000	.000	.000	.000
Atypical-Atypical:													
Single response B													
Predicted	.150	.167	.213	.166	.113	.073	.046	.028	.018	.011	.007	.004	.006
Observed	.150	.175	.173	.203	.118	.063	.035	.040	.018	.010	.008	.005	.005
Single response C													
Predicted	.125	.147	.240	.188	.123	.075	.044	.025	.014	.008	.004	.003	.003
Observed	.125	.155	.183	.248	.125	.075	.028	.033	.010	.010	.005	.000	.005
Combined responses BC													
Predicted	.090	.143	.232	.189	.130	.084	.052	.032	.019	.012	.007	.004	.006
Observed	.090	.150	.185	.238	.130	.083	.038	.043	.015	.013	.008	.003	.008
	Errors After First Success												
Typical-Typical:													
Single response B													
Predicted	.772	.155	.040	.011	.003	.001	.000						
Observed	.768	.168	.038	.020	.008	.000	.000						

TABLE 11 (Continued)

List condition and response analysis	Probability of k: 0	1	2	3	4	5	6	7	8	9	10	11	≥12
Single response C													
Predicted	.762	.165	.040	.010	.003	.001	.000						
Observed	.758	.170	.045	.020	.008	.000	.000						
Combined responses BC													
Predicted	.765	.155	.043	.013	.004	.001	.000						
Observed	.760	.163	.053	.018	.010	.000	.000						
<u>Typical-Atypical:</u>													
Single response B													
Predicted	.752	.155	.056	.020	.007	.003	.001						
Observed	.748	.168	.053	.030	.008	.003	.000						
Single response C													
Predicted	.739	.167	.059	.021	.008	.003	.001						
Observed	.743	.158	.065	.025	.008	.003	.000						
Combined responses BC													
Predicted	.742	.160	.059	.022	.008	.003	.001						
Observed	.735	.163	.063	.028	.013	.000	.000						
<u>Atypical-Typical:</u>													
Single response B													
Predicted	.834	.127	.028	.006	.001	.000	.000						
Observed	.833	.130	.028	.010	.000	.000	.000						
Single response C													
Predicted	.828	.121	.030	.008	.002	.001	.000						
Observed	.825	.115	.035	.015	.000	.000	.000						
Combined responses BC													
Predicted	.841	.113	.030	.008	.002	.001	.000						
Observed	.835	.125	.028	.008	.003	.000	.000						

TABLE 11 (Continued)

List condition and response analysis	0	1	2	3	4	5	6	7	8	9	10	11	>12
	Probability of *k*												
Atypical-Atypical:													
Single response B													
Predicted	.804	.106	.045	.020	.009	.004	.002						
Observed	.800	.123	.050	.020	.008	.000	.000						
Single response C													
Predicted	.820	.087	.043	.022	.011	.006	.003						
Observed	.803	.105	.053	.025	.005	.008	.003						
Combined responses BC													
Predicted	.832	.086	.041	.019	.009	.005	.002						
Observed	.815	.098	.058	.023	.003	.005	.000						
	Total Errors												
Typical-Typical:													
Single response B													
Predicted	.335	.239	.178	.110	.063	.035	.019	.010	.005	.003	.001	.001	.001
Observed	.335	.245	.168	.103	.073	.045	.010	.013	.008	.000	.003	.000	.000
Single response C													
Predicted	.335	.249	.180	.108	.060	.032	.017	.009	.005	.002	.001	.001	.001
Observed	.335	.248	.173	.118	.058	.043	.008	.010	.008	.000	.003	.000	.000
Combined responses BC													
Predicted	.290	.252	.191	.119	.068	.038	.020	.011	.006	.003	.002	.001	.001
Observed	.290	.255	.183	.115	.075	.050	.008	.015	.008	.000	.003	.000	.000
Typical-Atypical:													
Single response B													
Predicted	.170	.208	.195	.146	.100	.066	.042	.027	.017	.011	.007	.004	.006
Observed	.170	.220	.175	.153	.100	.070	.043	.018	.020	.010	.013	.005	.000
Single response C													
Predicted	.188	.230	.185	.133	.091	.060	.040	.026	.017	.011	.007	.005	.007
Observed	.188	.240	.165	.125	.108	.053	.055	.023	.013	.013	.010	.005	.000

TABLE 11 (Continued)

List condition and response analysis	Probability of k 0	1	2	3	4	5	6	7	8	9	10	11	≥12
Combined responses BC													
Predicted	.130	.200	.202	.156	.109	.073	.047	.030	.019	.012	.008	.005	.007
Observed	.130	.218	.180	.160	.100	.078	.055	.020	.023	.015	.013	.005	.005
Atypical-Typical:													
Single response B													
Predicted	.267	.204	.214	.139	.080	.044	.024	.013	.007	.004	.002	.001	.001
Observed	.268	.205	.183	.153	.075	.038	.023	.018	.008	.005	.000	.000	.000
Single response C													
Predicted	.255	.238	.209	.133	.076	.042	.022	.012	.006	.003	.002	.001	.001
Observed	.255	.245	.175	.165	.073	.043	.023	.018	.005	.000	.000	.000	.000
Combined responses BC													
Predicted	.208	.220	.225	.151	.089	.050	.027	.015	.008	.004	.002	.001	.001
Observed	.208	.228	.183	.193	.085	.045	.028	.020	.008	.005	.000	.000	.000
Atypical-Atypical:													
Single response B													
Predicted	.110	.162	.198	.167	.123	.085	.056	.037	.023	.015	.009	.006	.010
Observed	.110	.178	.200	.163	.115	.063	.050	.045	.038	.020	.008	.003	.010
Single response C													
Predicted	.105	.148	.204	.176	.130	.089	.058	.036	.022	.014	.008	.005	.007
Observed	.105	.155	.185	.198	.130	.073	.055	.038	.028	.018	.010	.003	.005
Combined responses BC													
Predicted	.073	.141	.204	.180	.136	.095	.063	.041	.026	.016	.010	.006	.015
Observed	.073	.148	.193	.203	.130	.073	.053	.053	.035	.023	.008	.003	.010

TABLE 11 (Continued)

List condition and response analysis	0	1	2	3	4	5	6	7	8	9	10	11	≥12
	Probability of *k*												
	Trial of Last Error												
<u>Typical-Typical:</u>													
Single response B													
Predicted	.335	.164	.135	.082	.046	.025	.013	.007	.004	.002	.001	.001	.001
Observed	.335	.173	.180	.058	.110	.078	.028	.013	.015	.008	.003	.000	.003
Single response C													
Predicted	.335	.167	.133	.078	.042	.022	.012	.006	.003	.002	.001	.000	.001
Observed	.335	.175	.180	.065	.103	.078	.028	.013	.010	.008	.003	.000	.003
Combined responses BC													
Predicted	.290	.180	.147	.088	.049	.026	.014	.007	.004	.002	.001	.001	.001
Observed	.290	.190	.185	.068	.110	.088	.030	.013	.015	.008	.003	.000	.003
<u>Typical-Atypical:</u>													
Single response B													
Predicted	.170	.189	.163	.116	.077	.049	.031	.020	.012	.008	.005	.003	.005
Observed	.170	.200	.163	.108	.115	.073	.048	.053	.025	.010	.013	.008	.018
Single response C													
Predicted	.188	.220	.148	.099	.066	.044	.029	.019	.012	.008	.005	.003	.007
Observed	.188	.230	.143	.090	.110	.053	.055	.055	.020	.020	.008	.010	.020
Combined responses BC													
Predicted	.130	.192	.171	.124	.083	.054	.035	.022	.014	.009	.006	.004	.070
Observed	.130	.208	.168	.113	.105	.083	.055	.053	.028	.018	.013	.010	.020
<u>Atypical-Typical:</u>													
Single response B													
Predicted	.267	.168	.187	.119	.067	.036	.020	.010	.006	.003	.002	.001	.001
Observed	.268	.173	.160	.193	.080	.043	.035	.018	.010	.010	.000	.000	.000
Single response C													
Predicted	.255	.199	.181	.109	.060	.032	.017	.009	.006	.003	.002	.001	.001
Observed	.255	.208	.173	.153	.078	.053	.038	.020	.010	.008	.008	.000	.000

TABLE 11 (Continued)

List condition and response analysis	Probability of k: 0	1	2	3	4	5	6	7	8	9	10	11	≥ 12
Combined responses BC													
Predicted	.208	.191	.204	.130	.074	.040	.022	.012	.006	.003	.002	.001	.001
Observed	.208	.200	.170	.190	.088	.050	.040	.025	.010	.010	.010	.000	.000
Atypical-Atypical:													
Single response B													
Predicted	.110	.140	.188	.153	.107	.070	.045	.028	.017	.011	.007	.004	.007
Observed	.110	.153	.188	.188	.095	.058	.053	.045	.040	.033	.018	.010	.015
Single response C													
Predicted	.105	.138	.209	.170	.115	.072	.043	.025	.014	.008	.004	.002	.003
Observed	.105	.145	.178	.200	.105	.073	.050	.038	.033	.035	.018	.008	.015
Combined responses BC													
Predicted	.073	.132	.207	.175	.124	.081	.051	.031	.019	.011	.007	.004	.006
Observed	.073	.133	.185	.208	.105	.075	.055	.048	.038	.035	.018	.010	.015

TABLE 12

Tests of Item Independence by List Condition for Experiment 2

List condition	Statistic $-2\log_e L_{\emptyset}$	$-2\log_e L_8$	$\chi^2(8)$
Typical-Typical	2552.17	2173.68	378.49
Typical-Atypical	2976.08	2512.54	463.54
Atypical-Typical	2450.91	2063.94	386.97
Atypical-Atypical	2910.85	2407.87	502.98

NOTE. The critical value that the test statistic in the third column must exceed in order to reject the null hypothesis of item independence is 15.51 at $p < .05$.

based on these results as well as those from Experiment 1, that items in semantically unrelated and related associative clusters are represented in a single memory structure.

Insert Table 13 about here

Acquisition processes: Strength versus all-or-none. The results of the overall goodness-of-fit analyses revealed that the two-stage Markov model provided both a necessary and a sufficient account of the data. That is, consistent with the all-or-none or unitary trace view, acquisition consisted of exactly three discrete performance states and transitions between states (i.e., learning) were all-or-none. More important, this finding is inconsistent with the strength assumption in the schema version of the multiple trace theory.

To investigate this issue in a more detailed manner, fine-grained statistical analyses of within-state stationarity were conducted. Specifically, the predicted and observed distributions of the same two intermediate State P statistics used in Experiment 1 were compared. The distributions of these statistics are shown in Figures 8-11 for the combined responses BC analyses and in Table 14 for the single response B and single response C analyses. Out of a total of 24 Kolmogorov-Smirnov tests (4 lists x 3 response analyses x 2 statistics) not one null hypothesis rejection was obtained ($p < .05$). That is, the observed distributions of these statistics were accurately predicted from the two-stage model. These data are inconsistent with any model that assumes that response probability varies continuously rather than discretely across trials.

TABLE 13

Tests of Item Dependence by Hypothesis and List Condition for Experiment 2.

Hypothesis and List condition	Statistic $-2\log_e L_8$	$-2\log_e L_{16}$	$\chi^2(8)$
P(B) = P(C):			
Typical-Typical	4204.76	4204.21	0.55
Typical-Atypical	4943.73	4937.87	5.86
Atypical-Typical	4067.84	4062.62	5.22
Atypical-Atypical	4943.25	4933.33	9.92
P(B) = P(B ∩ C):			
Typical-Typical	4283.73	4281.14	2.59
Typical-Atypical	4994.72	4989.46	5.26
Atypical-Typical	4091.68	4085.09	6.59
Atypical-Atypical	4915.08	4903.41	11.67
P(C) = P(B ∩ C):			
Typical-Typical	4281.37	4270.43	10.94
Typical-Atypical	4982.60	4973.49	9.11
Atypical-Typical	4111.93	4105.41	6.52
Atypical-Atypical	4852.05	4845.66	6.39

NOTE. The critical value that the test statistic in the third column must exceed in order to reject the null hypothesis of item dependence is 15.51 at $p < .05$.

Insert Figures 8-11 and Table 14 about here

Taken together, the results of the goodness-of-fit analyses for both experiments demonstrate a high degree of correspondence between the two-stage model and the global as well as fine-grained aspects of the data. Contrary to the strength assumption in the schema version of the multiple trace theory, neither goodness-of-fit nor within-state stationarity fluctuated as a function of variation in the degree of interitem relatedness. That is, the two-stage model provided an equally adequate account of the acquisition data from semantically unrelated list conditions (Experiment 1) and from lists where associative clusters varied in their degree of category membership (Experiment 2).

Loci of preexperimental organization effects: The manipulation of typicality in the presence of category membership. Three issues will be dealt with in the following sections. The first issue is whether or not the manipulation of preexperimental semantic relatedness affected acquisition. This issue can be broken down into two questions, (a) the manner in which the presence versus the absence of preexperimental knowledge affected acquisition, and (b) the manner in which variation in the degree of category membership affected acquisition. Question (a) asks about the difference between learning semantically unrelated associative clusters (Experiment 1) and learning semantically related associative clusters (Experiment 2). This question can be answered by conducting both omnibus and parameter-specific comparisons between corresponding list conditions

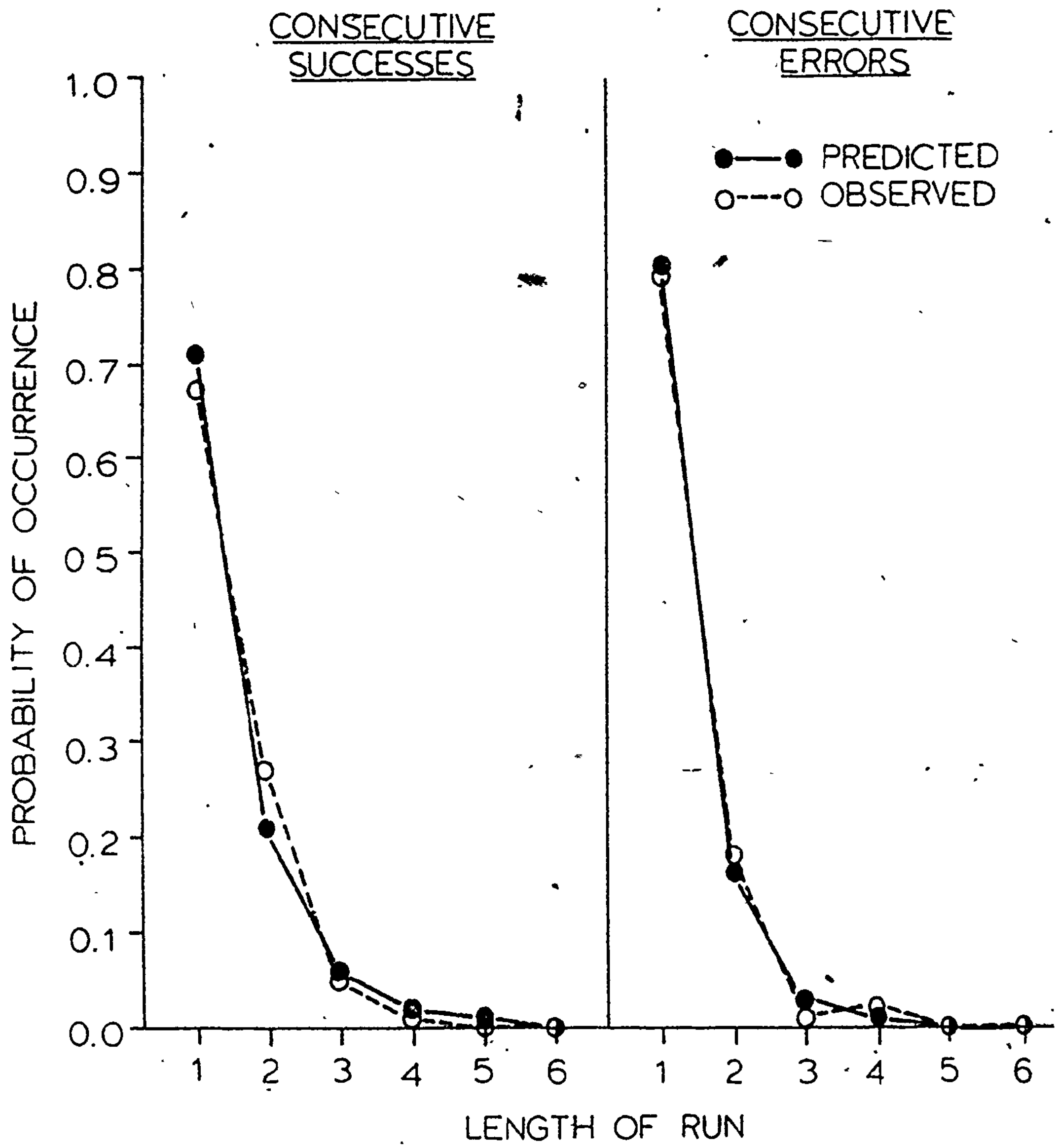

Figure 8. Predicted and Observed Distributions of Two Intermediate State Statistics, Length of Success Runs Before the Last Error (Left Panel) and Length of Error Runs After the First Success (Right Panel), for the Combined Response Data from the Typical-Typical Triad Condition in Experiment 2.

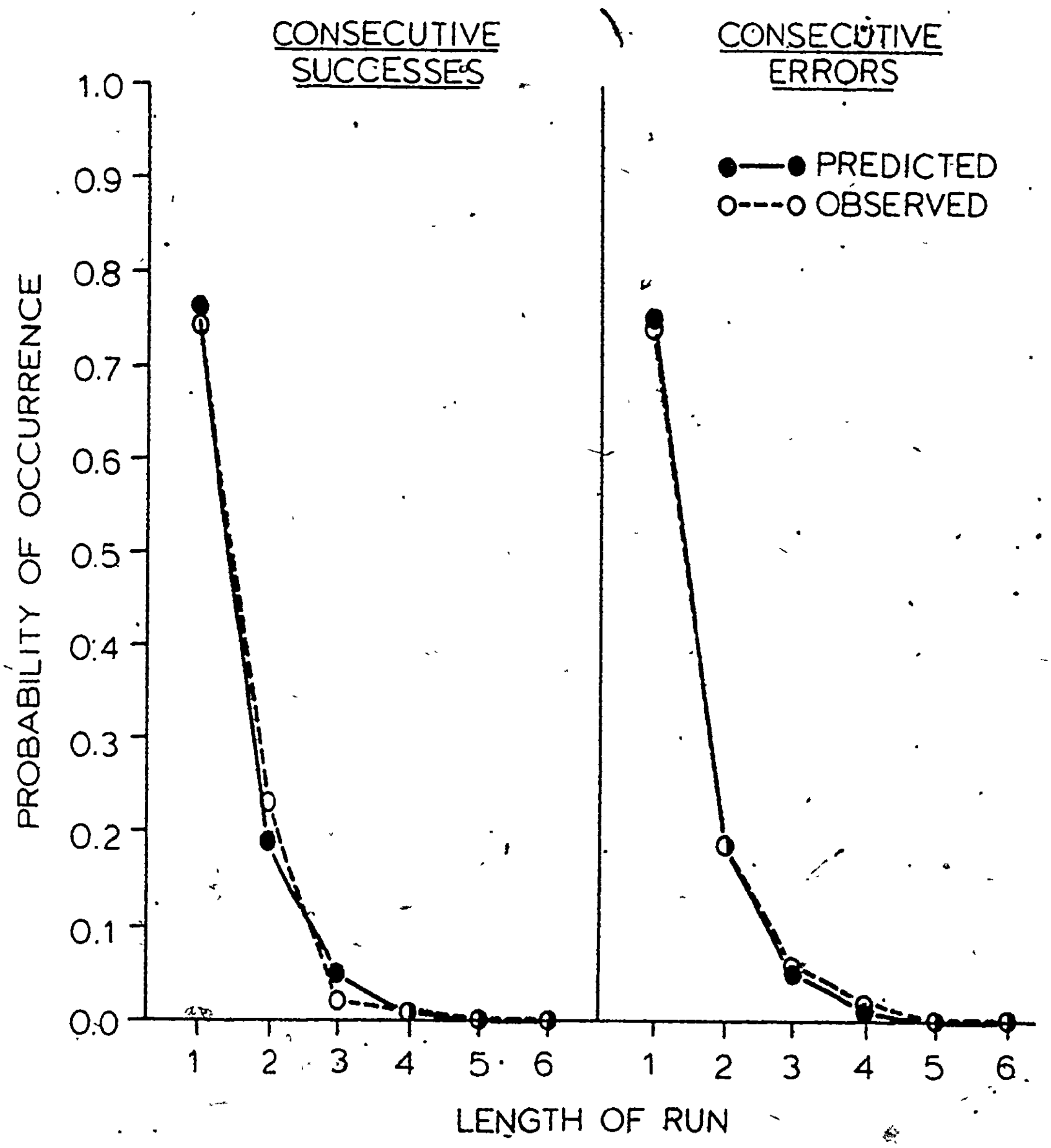

Figure 9. Predicted and Observed Distributions of Two Intermediate State Statistics, Length of Success Runs Before the Last Error (Left Panel) and Length of Error Runs After the First Success (Right Panel), for the Combined Response Data from the Typical-Atypical Triad Condition in Experiment 2.

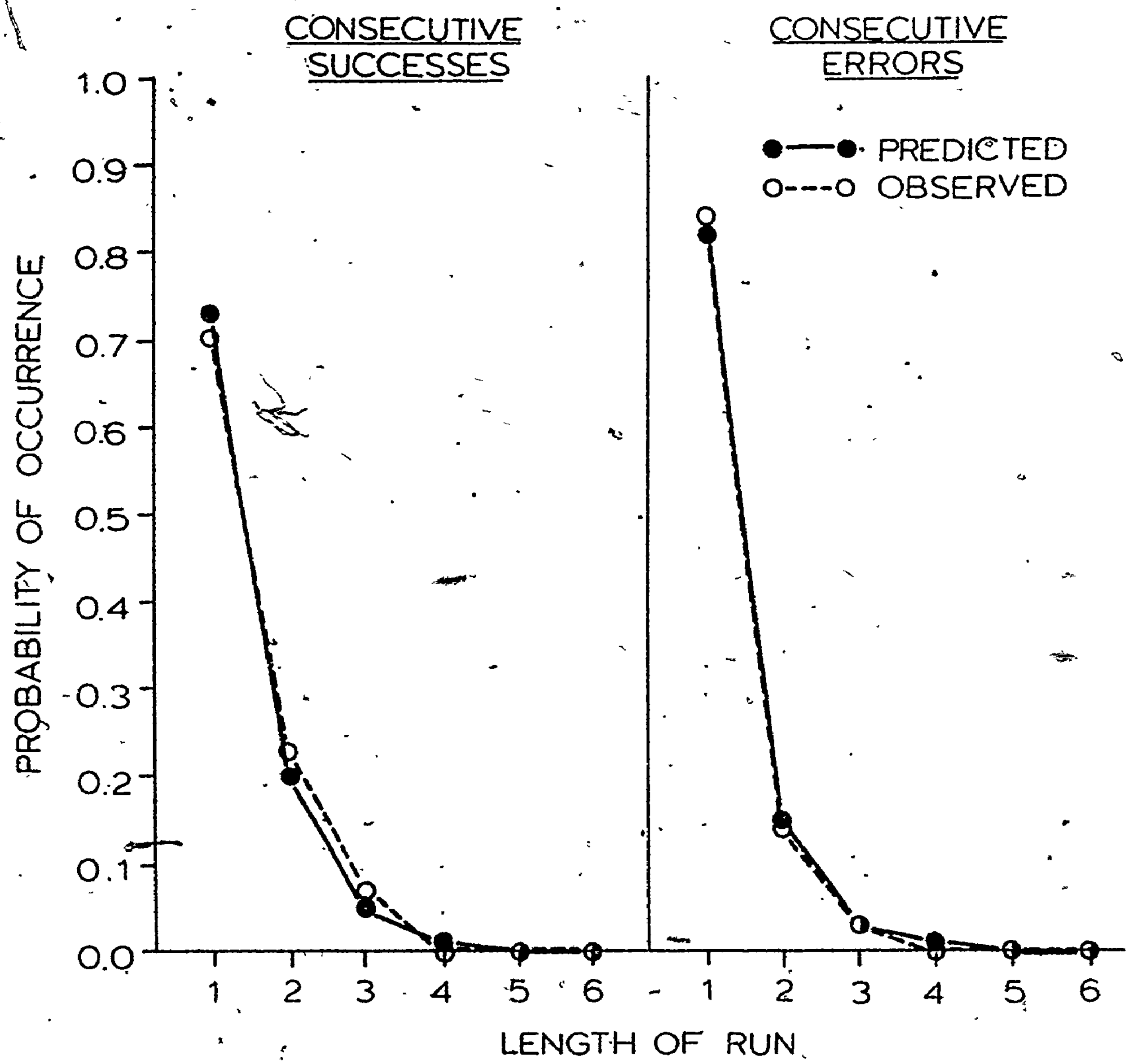

Figure 10. Predicted and Observed Distributions of Two Intermediate State Statistics, Length of Success Runs Before the Last Error (Left Panel) and Length of Error Runs After the First Success (Right Panel), for the Combined Response Data from the Atypical-Typical Triad Condition in Experiment 2.

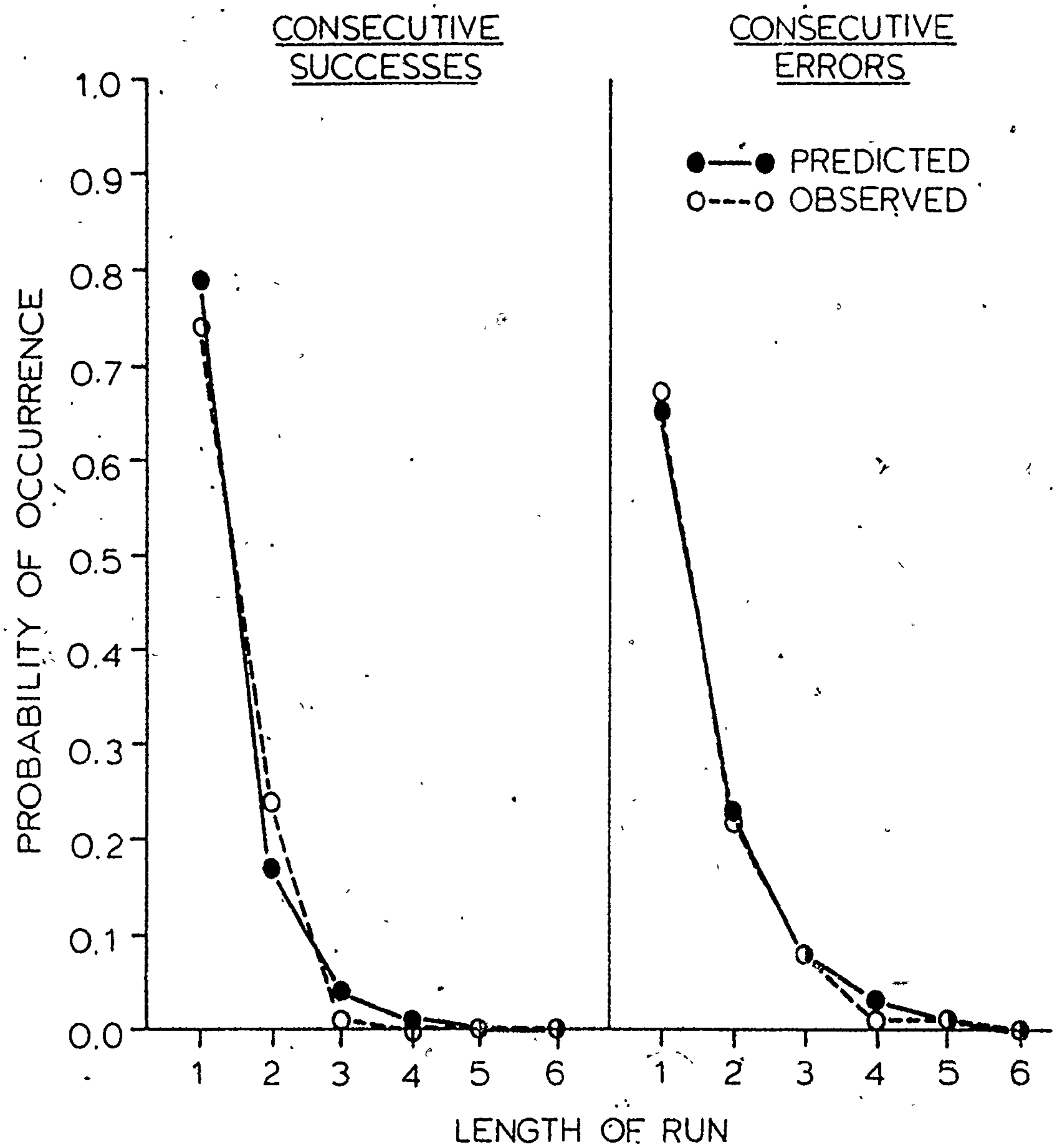

Figure 11. Predicted and Observed Distributions of Two Intermediate State Statistics, Length of Success Runs Before the Last Error (Left Panel) and Length of Error Runs After the First Success (Right Panel), for the Combined Response Data from the Atypical-Atypical Triad Condition in Experiment 2.

TABLE 14

Predicted and Observed Distributions of Two Intermediate-State Statistics of the Individual Response Data by List Condition for Experiment 2.

List condition and response analysis	1	2	3	4	5	6	7	>8
	Probability of *k*							
	Length of Consecutive Success Runs before the Last Error							
Typical-Typical								
Single response B								
Predicted	.686	.216	.068	.021	.007	.002	.001	.000
Observed	.670	.239	.055	.037	.000	.000	.000	.000
Single response C								
Predicted	.695	.212	.065	.020	.006	.002	.001	.000
Observed	.636	.307	.053	.009	.000	.000	.000	.000
Typical-Atypical								
Single response B								
Predicted	.742	.191	.049	.013	.003	.001	.000	.000
Observed	.712	.237	.042	.009	.000	.000	.000	.000
Single response C								
Predicted	.718	.203	.057	.016	.005	.001	.000	.000
Observed	.656	.303	.033	.008	.000	.000	.000	.000
Atypical-Typical								
Single response B								
Predicted	.686	.216	.068	.021	.007	.002	.001	.000
Observed	.667	.236	.069	.028	.000	.000	.000	.000
Single response C								
Predicted	.703	.209	.062	.019	.006	.002	.001	.000
Observed	.649	.273	.078	.000	.000	.000	.000	.000
Atypical-Atypical								
Single response B								
Predicted	.767	.179	.042	.010	.002	.001	.000	.000
Observed	.717	.263	.020	.000	.000	.000	.000	.000
Single response C								
Predicted	.752	.187	.046	.012	.003	.001	.000	.000
Observed	.708	.250	.042	.000	.000	.000	.000	.000

TABLE 14 (Continued)

List condition and response analysis	Probability of k: 1	2	3	4	5	6	7	≥8
	Length of Consecutive Error Runs after the First Success							
Typical-Typical								
Single response B								
Predicted	.814	.152	.028	.005	.001	.000	.000	.000
Observed	.807	.165	.018	.009	.000	.000	.000	.000
Single response C								
Predicted	.828	.143	.025	.004	.001	.000	.000	.000
Observed	.807	.175	.009	.009	.000	.000	.000	.000
Typical-Atypical								
Single response B								
Predicted	.747	.189	.048	.012	.003	.001	.000	.000
Observed	.754	.161	.068	.017	.000	.000	.000	.000
Single response C								
Predicted	.749	.188	.047	.012	.003	.001	.000	.000
Observed	.738	.205	.049	.008	.000	.000	.000	.000
Atypical-Typical								
Single response B								
Predicted	.826	.144	.025	.004	.001	.000	.000	.000
Observed	.833	.139	.028	.000	.000	.000	.000	.000
Single response C								
Predicted	.843	.132	.021	.003	.001	.000	.000	.000
Observed	.844	.130	.026	.000	.000	.000	.000	.000
Atypical-Atypical								
Single response B								
Predicted	.687	.215	.067	.021	.007	.002	.001	.000
Observed	.697	.212	.071	.010	.010	.000	.000	.000
Single response C								
Predicted	.651	.227	.079	.028	.010	.003	.001	.000
Observed	.646	.250	.083	.010	.010	.000	.000	.000

in the two experiments. Question (b) asks about the difference between learning typical and atypical category members. This question can be answered by conducting omnibus and parameter-specific comparisons between list conditions in Experiment 2.

The second issue is whether the probability of learning an item increased following a success or following an error when subjects recalled related associative clusters rather than unrelated clusters. This question can be answered by testing the two hypotheses, $\underline{c} > 0$ and $\underline{d} > 0$. Finally, the third issue is whether the same list-invariant ordinal relationships that were observed between learning and performance parameters in Experiment 1 were also observed in Experiment 2. This question can be answered by testing the two parameter-invariance hypotheses, $\{\underline{a}', \underline{a}\} \leq \{\underline{b}', \underline{b}, \underline{d}\}$ and $\underline{g} \geq \underline{h}$.

Each of these issues will be dealt with, in turn, below. Because interest in these theoretical questions focuses on the learning of associative clusters, all subsequent analyses were conducted on the combined responses BC data. Prior to examining each of these hypotheses, however, it was necessary to obtain evidence that: (1) At a global level, the parameter values differed between the list conditions of Experiments 1 and 2; (2) At a global level, the parameter values differed between list conditions in Experiment 2; (3) An identifying restriction can be located for the data from Experiment 2 in order to obtain numerical estimates of the theoretical parameters; and (4) The identifying restriction used in Experiment 2 is the same as the one used in Experiment 1.

To determine whether there was global statistical evidence that parameter values differed between lists conditions in the two experiments as a whole, an experimentwise likelihood ratio test was used. The null hypothesis that there were no differences in learning as a function of manipulating the presence versus the absence of preexperimental knowledge was tested by simply: (1) multiplying the separate likelihoods of each list condition together to form a single likelihood value L_{64} for both experiments; and (2) obtaining the joint likelihood value L_8 of the same data by pooling the data from both experiments and minimizing the likelihood function. The appropriate likelihood ratio test consists of taking twice the negative natural log of L_8 divided by L_{64}. The distribution of this test statistic is $\chi^2(56)$ and the numerical result was $\chi^2(56) = 2150.43$, $p < .0001$. This result indicates that there were large differences in learning semantically unrelated versus semantically related associative clusters.

The same likelihood ratio procedure was used to determine whether there was global statistical evidence that parameter values differed between list conditions in the second experiment. The null hypothesis that there were no differences in learning as a function of manipulating typicality in the presence of category membership was tested by simply: (1) multiplying the separate likelihoods of each condition together to form a single likelihood value L_{32} for the second experiment; and (2) obtaining the joint likelihood value L_8 of the same data by pooling the data from all of the list conditions together prior to minimizing the likelihood function. The appropriate likelihood ratio test consists of taking twice the negative natural log of L_8 divided by L_{32}. The distribution of this test statistic is $\chi^2(24)$ and the numerical result was $\chi^2(24) = 238.62$, $p < .0001$. This result indicates that there were substantial differences in learning between the different list conditions in Experiment 2. That is, typicality, when manipulated in the presence of category membership, had a strong impact on the acquisition of associative clusters.

Identification and estimation of the theoretical parameters of the two-stage model. Before considering the parameter-specific effects of manipulating preexperimental semantic relatedness, it is necessary to locate an identifying restriction. The problem here is slightly more complicated than in Experiment 1. Recall that in order to evaluate the various hypotheses about the influence of preexperimental knowledge on acquisition, both between-condition and between-experiment tests must be conducted. Because spurious differences can result when conditions having different identifying restrictions are being compared (see Brainerd et al., 1982), it is necessary to find an identifying restriction that is valid across all list conditions in both experiments.

The obvious starting point was to test the identifying restriction $\underline{r} = \underline{c} = 0$, and $1 - \underline{a}' \geq \underline{a}'\underline{b}'(1-\underline{a}'\underline{b}')$ that was found to be valid for each of the four list conditions in Experiment 1. The validity of this restriction for the four list conditions in Experiment 2 was evaluated by obtaining the likelihood of the data when all 8 identifiable parameters were free to vary, L_8, and by obtaining the likelihood of the same data subject to the identifying restriction, L_7. The appropriate likelihood ratio test of the null hypothesis that this restriction did not significantly reduce the likelihood of the data consisted of taking twice the negative natural log of L_7 divided by L_8. The distribution of test statistic is $\chi^2(1)$

and the numerical result for all list conditions was $\chi^2(1) = 0.00$, $p > .05$. This result indicates that this restriction was valid for all four list conditions in Experiment 2. More importantly, this finding indicates that a single identifying restriction was valid for all eight list conditions in the experiments reported here. The numerical values of the storage parameters, retrieval-learning parameters, and retrieval-performance parameters appear in the last 4 rows of Table 8.

Between-experiment comparisons. Because the same identifying restriction was valid for all list conditions in Experiments 1 and 2, the question of how acquisition was affected by the absence versus the presence of preexperimental knowledge can be investigated. Recall that both the multiple and unitary trace theories make the rather trivial prediction that semantically related clusters should be easier to learn than semantically unrelated clusters. The more interesting question concerns the precise locus of these effects at both a stage-specific and parameter-specific level. Three loci will be considered, namely, first-stage storage parameters (a' and a), second-stage retrieval-learning parameters (b', b, and d) and second-stage retrieval-performance parameters ($1-e$, g, and h).

Concerning the first-stage learning parameters a' and a, both versions of the unitary trace theory and the schema version of the multiple trace theory agree that the values of these parameters should be higher in conditions where items within a cluster are related (Experiment 2) than in conditions where they are unrelated (Experiment 1). This is because relations between concepts are viewed as important both during storage and during retrieval. The HAM version of the multiple trace theory, on the other hand, makes the prediction that interitem semantic relatedness should not affect the first stage of acquisition. This is because the first stage simply involves storing (i.e., activating and tagging) the separate cue and target concepts in memory, a process that does not depend on whether the concepts are related or unrelated.

Concerning the second-stage learning parameters b', b and d, both versions of the multiple and unitary trace theories agree that the values of these parameters should be higher in related than unrelated list conditions. Finally, only one model, namely, the modified storage-retrieval version of the unitary trace theory) makes predictions about the second-stage performance parameters 1-e, g, and h. Specifically, this model predicts that the values of these parameters should remain invariant across manipulations of interitem relatedness. This is because second-stage performance parameters, which are assumed to be measures of heuristic retrieval, should be insensitive to the nature of the material being studied (i.e., they are context-free retrieval operations).

There are three steps involved in testing these parameter-specific hypotheses. First, it is necessary to determine whether the values of the parameters differed between the two experiments as a whole. As reported earlier, the between-experiment omnibus test was significant, indicating that at a global level the parameter values differed between Experiments 1 and 2.

Second, it is necessary to determine which pairs of conditions differed between the two experiments. Rather than consider all possible pairwise tests, comparisons are typically conducted only on those pairs of conditions for which possible parameter differences are of interest. In the present section, interest centers on differences between learning related and unrelated associative clusters. Because the same cue and target items were used in both experiments, the most precise way of investigating this issue is to compare each of the list conditions in Experiment 1 (i.e., TT_1, TA_1, AT_1, and AA_1) with its corresponding list condtiion in Experiment 2 (i.e., TT_2, TA_2, AT_2, and AA_2). These conditionwise tests, which can be considered an analog of the t test, are similar to the omnibus experimentwise tests. Quite simply, the conditionwise test evaluates the null hypothesis that the parameter values do not differ between any pair of conditions. This test is computed, subject to the identifying restriction, by: (1) multiplying the separate likelihoods of both list conditions together to form a single likelihood value L_{14}; and (2) obtaining the joint likelihood value L_7 of the same data by pooling the data from both list conditions together prior to minimizing the likelihood function. The appropriate likelihood ratio test consists of taking twice the negative natural log of L_7 divided by L_{14} (see Brainerd et al., 1982, Equations 46a-46b). The distribution of this test statistic is $\chi^2(7)$. The numerical results were $\chi^2(7) = 402.70$, $p < .0001$ (TT_1 versus TT_2), $\chi^2(7) = 305.24$, $p < .0001$ (TA_1 versus TA_2), $\chi^2(7) = 442.83$, $p < .0001$ (AT_1 versus AT_2), and $\chi^2(7) = 267.68$, $p < .0001$ (AA_1 versus AA_2).

Given that all of the conditionwise tests were significant, the third and final step is to compute parameterwise tests for each of the eight parameters in Table 8. The parameterwise test evaluates the null hypothesis that the value of some parameter (say b) does not differ between two conditions. This test is computed, subject to the identifying restriction, by: (1) multiplying the separate likelihoods of both conditions together to form a single likelihood value L_{14}; and (2) obtaining the joint likelihood L_{13} of the same data by minimizing both sets of data simultaneously subject to the additional restriction that the parameter of interest (say *b*) takes on the same numerical value in each condition. The appropriate likelihood ratio test consists of taking twice the negative natural log of L_{13} divided by L_{14} (see Brainerd et al., 1982, Equations 47a-47b). The distribution of this test statistic is $\chi^2(1)$. Because the total number of parameterwise tests was 32 (4 conditionwise comparisons x 8 parameters), detailed numerical results will be omitted. A summary of the significant between-experiment parameter differences will be presented instead.

Concerning the first-stage storage parameters, an examination of the first two columns of Table 8 shows that the values of *a'* and *a* were higher in the second experiment than in the first experiment for all of the conditionwise list comparisons. Parameterwise tests revealed that these differences were significant ($p < .01$). The average between-experiment difference for the first-stage learning parameters was .22, where the mean difference between the values of *a'* was .19 and the mean difference between the values of *a* was .26. Contrary to the HAM version of the multiple trace theory, this result indicates that the ease of storing items in memory depends on whether or not those items are preexperimentally related. That is, consistent with both versions of the unitary trace theory and the schema version of the multiple trace theory, the values of *a'* and *a* were higher (i.e., storage of information was facilitated) in conditions where items within a cluster were related (Experiment 2) than in conditions where they were unrelated (Experiment 1).

Concerning the second-stage retrieval-learning parameters, an examination of the third-fifth columns of Table 8 shows that the values of $\underline{b}'$, $\underline{b}$, and $\underline{d}$ were higher in the second experiment than in the first experiment for all of the conditionwise list comparisons. Parameterwise tests revealed that all of these differences were significant ($\underline{p} < .01$) with the exception of $\underline{b}'$ and $\underline{d}$ for the AA_1 versus AA_2 comparisons. The average between-experiment difference for the second-stage retrieval-learning parameters was .29, where the mean difference between the values of $\underline{b}'$ was .29, between the values of $\underline{b}$ was .42, and between the values of $\underline{d}$ was .16. This result is consistent with both versions of the multiple and unitary trace theories.

Finally, concerning the second-stage retrieval-performance parameters, an examination of the sixth-eighth columns of Table 8 shows that: (1) The values of 1-$\underline{e}$ (the probability of a success on the first trial in the intermediate State P given the item entered P on any trial after the first trial) were higher in the second experiment than in the first experiment; and (2) The values of $\underline{g}$ and $\underline{h}$ remained invariant across experiments. Parameterwise tests revealed that the values of 1-$\underline{e}$ were significantly higher in the second experiment ($\underline{p} < .05$) except for the AA_1 versus AA_2 comparison. In addition, these tests revealed that the values of $\underline{g}$ and $\underline{h}$ did not differ significantly between experiments with the one exception of $\underline{g}$ for the AA_1 versus AA_2 comparison ($\underline{p} < .05$). This latter result is consistent with the modified storage-retrieval version of the unitary trace theory.

Taken together, the results of the stages-of-learning analysis indicate that the addition of preexperimental knowledge has different effects on acquisition depending on whether one is talking about learning or about performance. With regard to learning, both the encoding and storage of information (first-stage learning parameters $\underline{a}'$ and $\underline{a}$) and learning to retrieve information from stored traces (second-stage learning parameters, $\underline{b}'$, $\underline{b}$, and $\underline{d}$) was easier when items within a cluster were categorically related than when they were unrelated. Moreover, there was no indication that preexperimental knowledge interacted with stage of learning as the overall magnitude of these effects were approximately equivalent across states (State U = .22 versus State P = .29). These results, while anticipated by most theories, rule out models such as HAM that claim that early learning consists of simply activating and tagging isolated conceptual units in memory.

With regard to performance, two interesting trends emerged. First, given that an item made a transition from the first to the second stage of learning on any trial after the first study trial, the probability of making a correct response on the first trial in the second stage ($1-\underline{e}$) was greater in Experiment 2 than in Experiment 1. That is, initial performance in the second stage of learning was better when items were semantically related than when they were unrelated. Second, performance on all trials after the first trial in the second acquisition stage (parameters $\underline{g}$ and $\underline{h}$) remained invariant across experiments. That is, the majority of second-stage performance did not vary as a function of whether subjects were learning related or unrelated associative clusters.

This global pattern of results is generally consistent with the modified storage-retrieval version of the unitary trace theory. Specifically, these findings are in agreement with the proposal that second-stage performance is mediated by a context-free heuristic retrieval system. The one exception was that in three out of four comparisons, performance on the first trial in the second stage was better when items were related than when they were unrelated. This finding does not, however, constitute serious evidence against the modified storage-retrieval interpretation of second-stage learning. Because there is very little data available to estimate the value of _e_ (only Trial 1 in the second stage), numerical estimates of this parameter are typically unstable. Furthermore, the extent of the available data base decreases as the values of _b'_ and _b_ (i.e., as the number of items that skip the second stage) increase. In general, as the stability of a parameter estimate decreases, the likelihood of obtaining a spurious between-condition difference increases. The extent to which it is theoretically meaningful to interpret any between-condition difference in the parameter _e_ is, therefore, limited. This is particularly true here in that comparisons were conducted between conditions that differed substantially in their _b'_ and _b_ values (see Table 8). It is not unreasonable to conclude, therefore, that the second-stage performance data, in particular the behavior of the parameters _g_ and _h_, are consistent with a modified storage-retrieval view.

Between-condition comparisons. The next question is whether or not variation in the degree of category membership affected acquistion. Recall that it is only the HAM version of the multiple trace theory that predicts that learning should not vary as a function of item typicality. Both versions of the unitary trace theory and the schema version of the multiple trace theory, on the other hand, predict that acquisition should be affected by variation in the overall degree of intracluster similarity. Specifically, these models predict that the overall ease of list acquisition should vary in the following manner: TT $\rightarrow$ AT $\rightarrow$ TA $\rightarrow$ AA.

The HAM position is clearly an untenable one given that the experimentwise omnibus test reported earlier revealed significant differences between the list conditions in Experiment 2. Inspection of Table 8 shows that, in general, changes in the parameter values for Experiment 2 correspond to the prediction that list difficulty should follow the pattern TT $\rightarrow$ AT $\rightarrow$ TA $\rightarrow$ AA. The numerical results for the three relevant conditionwise tests were $X^2(7) = 32.60$, $p < .001$ (TT versus AT), $X^2(7) = 38.81$, $p < .001$ (AT versus TA), and $X^2(7) = 42.76$, $p < .001$ (TA versus AA). Parameterwise tests for each of the list comparisons revealed that first-stage learning parameters, second-stage learning parameters, and one second-stage performance parameter were affected by variation in the degree of intracluster similarity. Specifically, $X^2(1)$ tests showed that: (1) a' and $1-e$ were significantly higher, but b' was significantly lower, in the TT list condition than in the AT list condition; (2) a', a, d, and $1-e$ were significantly higher in the AT list condition than in the TA list condition; and (3) a', b, d, and $1-e$ were significantly higher in the AT list condition than in the AA list condition.

These results confirm the global hypothesis that acquisition is facilitated when intracluster similarity increases. A more interesting question, however, concerns the stage-specific and parameter-specific effects of varying typicality on the cue versus the target side of associative clusters. Three hypotheses will be considered, each of which make different predictions about the loci of these effects in terms of first-stage storage parameters (<u>a</u>' and <u>a</u>), second-stage retrieval-learning parameters (<u>b</u>', <u>b</u>, and <u>d</u>), and second-stage retrieval-performance parameters (1-<u>e</u>, <u>g</u>, and <u>h</u>).

Recall that in the schema version of the multiple trace theory, the entire process of acquisition is said to be affected by the overall degree of relatedness between all of the concepts within a cluster. According to this model, both first-stage and second-stage learning parameters (<u>a</u>', <u>a</u>, <u>b</u>', <u>b</u> and <u>d</u>) should increase as a function of both cue and target typicality. Similarly, the modified storage-retrieval version of the unitary trace theory predicts that the parameters that measure storage and retrieval learning should increase as a function of both cue and target typicality. In addition, this model predicts that the second-stage retrieval-performance parameters (1-<u>e</u>, <u>g</u>, and <u>h</u>) should remain invariant across manipualtions of cue and target typicality. Finally, both the encoding-specificity and the early storage-retrieval versions of the unitary trace theory predict that first-stage parameters (<u>a</u>' and <u>a</u>) should be affected by both cue and target typicality. However, because retrieval is viewed as a cue dependent process, cue typicality should have a more powerful effect on second-stage learning parameters (<u>b</u>', <u>b</u>, and <u>d</u>) than target typicality.

Two conditionwise tests are relevant in assessing the effects of cue typicality, namely, TT versus AT and TA versus AA. The numerical results for the cue comparisons were $\chi^2(7) = 32.60$, $p < .001$ (TT versus AT) and $\chi^2(7) = 42.76$, $p < .001$ (TA versus AA). Parameterwise tests indicated that the advantage of typical over atypical cues was limited to one first-stage learning parameter, second-stage learning parameters, and one second-stage performance parameter. Specifically, $\chi^2(1)$ tests showed that: (1) a' and $1-e$ were significantly higher, but b' was significantly lower, in the TT list condition than in AT list condition; and (2) a', b, d, and $1-e$ were significantly higher in the TA list condition than in the AA list condition. The two performance parameters, g and h, did not differ for either comparison. Notice that both the magnitude and the direction of the cue effects depended somewhat on target difficulty. The overall advantage of typical cues, regardless of target type, was in the first-stage learning parameter a' and the second-stage performance parameter $1-e$. However, when the targets were also typical the second-stage learning parameter b' was negatively affected, but when the targets were atypical the second-stage learning parameters b and d were positively affected.

Two conditionwise tests are relevant in assessing the effects of target typicality, namely, TT versus TA and AT versus AA. The numerical results for the target comparisons were $\chi^2(7) = 92.33$, $p < .001$ (TT versus TA), and $\chi^2(7) = 84.27$, $p < .001$ (AT versus AA). Parameterwise tests indicated that the advantage of typical over atypical targets was in both first-stage learning parameters, second-stage learning parameters, and one second-stage performance parameter. Specifically, $\chi^2(1)$ tests showed that: (1) a' and a were significantly higher, but b' was significantly lower, in the TT list condition than in the TA list condition; and (2) a', a, b, d and 1-e were significantly higher in the AT list condition than in the AA list condition. Again, the two performance parameters, g and h, did not differ for either comparison. This pattern of results is similar to that obtained for cue typicality. The one difference was that target typicality not only had an overall positive effect on the first-stage learning parameter a', but had an additional positive effect on the first-stage learning parameter a. Again, an interaction between cue and target typicality was observed as both the magnitude and the direction of target effects depended somewhat on cue difficulty. That is, when the cues were typical the second-stage learning parameter b' was negatively affected, but when the cues were atypical the second-stage learning parameters b and d were positively affected.

Summing up, first-stage and second-stage learning parameters were affected by both cue and target typicality. This is consistent with all but one of the three models presented above, namely, the encoding-specificity and early storage-retrieval versions of the unitary trace theory. Recall that according to this model, cue difficulty should have had a larger impact on second-stage learning than target difficulty. Clearly, this was not the case as the results showed that both cue and target difficulty had equivalent effects on second-stage learning parameters. In fact, the only difference between manipulating typicality on the cue versus the target side was not on the second-stage parameters, but rather on the first-stage learning parameter a. Finally, consistent with the predictions of the modified storage-retrieval model, both second-stage performance parameters, g and h, remained invariant across manipulations of cue and target typicality. While these findings favor a unitary trace interpretation, they cannot be interpreted as evidence against a multiple trace interpretation. This is because predictions about the behavior of second-stage performance parameters were absent in the schema model.

Within-condition comparisons. The last two questions have to do with parameter-specific, within-condition hypotheses. The first question is whether the probability of learning an item increased following a success, as proposed in the schema model, or following an error, as proposed in the unitary trace model. Recall that these hypotheses imply the parameter-specific predictions $c > 0$ and $d > 0$, respectively. As in the first experiment, the hypothesis $c > 0$ was tested in conjunction with the evaluation of the validity of the identifying restriction for the list conditions in Experiment 2. Because $c = 0$ was a component of the identifying restriction that was found to be acceptable for all of the list conditions in Experiment 2, it follows that c could not have been greater than zero for these data. Contrary to the multiple trace prediction, then, the probability of learning an item did not increase as a function of successful retrieval attempts.

On the other hand, the prediction $d > 0$ was found to be true for these data. As can be seen in Table 8 the value of d is substantially greater than zero for all of the list conditions in the second experiment. When tested statistically using standard likelihood ratio procedures, the null hypothesis that $d = 0$ was rejected in every case, $\chi^2(1)$ 10.00, $p < .05$. Thus, consistent with the unitary trace model, the probability of learning increased following an error. It is interesting to note that this effect, namely, that the probability of learning increased following an error rather than following a success, was obtained regardless of the nature of the material being memorized. That is, this finding was observed independent of whether or not items within a cluster were related (Experiment 2) or unrelated (Experiment 1) and regardless of variation in the degree of intracluster similarity (Experiment 2).

The second and final question also concerns list-invariant effects. Specifically, concern is with whether or not the parameter-invariant relationships observed in the first experiment, namely, (1) $\{\underline{a}', \underline{a}\} \leqq \{\underline{b}', \underline{b}, \underline{d}\}$, and (2) $\underline{g} \geqq \underline{h}$, were also observed in the second experiment. Concerning 1, likelihood ratio tests revealed that $\{\underline{a}', \underline{a}\} < \{\underline{b}', \underline{b}, \underline{d}\}$ for the TT, TA, and AT list conditions and $\underline{a}' < \{\underline{b}', \underline{b}, \underline{d}\}$, $\underline{a} < \{\underline{b}', \underline{d}\}$ for the AA list condition. These findings are in general agreement with those obtained in the first experiment and indicate that first-stage learning was more difficult than second-stage learning. More importantly, these results suggest that the most difficult component of learning associative clusters, whether related or unrelated, was storing the information in memory, not in learning to retrieve that information from memory.

Concerning 2, inspection of Table 8 suggests that the $\underline{g} \geqq \underline{h}$ pattern observed in Experiment 1 was also obtained in Experiment 2. While this trend exists, likelihood ratio tests failed to reveal any significant difference between the values of $\underline{g}$ and $\underline{h}$ in the second experiment. That is, the probability of a success in the second-stage of learning did not differ as a function of whether it was preceded by another success or by an error.

GENERAL DISCUSSION

The purpose of the present investigation was to examine two fundamental aspects of associative memory, namely, the structure of associative memory representations and the nature of the processes involved in their acquisition. These issues were considered in the context of two influential theories of memory, namely, the multiple trace theory and the unitary trace theory. The overall pattern of results from the two experiments reported here provided overwhelming support for the unitary trace hypothesis that (1) the memory representation for associative clusters consisted of holistic, not independent structures, and (2) the acquisition of these unitary traces consisted of a two-stage discrete event rather than a single continuous process. More importantly, these findings were invariant with respect to manipulations of preexperimental semantic relatedness (Experiments 1 versus 2) including variation in the degree of intracluster similarity (Experiment 2). Discussion of these data will be grouped under three headings, namely, results that are relevant to trace structure, results that are relevant to trace acquisition, and results that are relevant to the manipulation of preexperimental intracluster similarity.

Trace structure: Independence versus dependence.

The findings of both experiments were consistent with respect to the question of trace structure. Regardless of whether items within a cluster were from the same semantic category (Experiment 2) or from different semantic categories (Experiment 1), the results clearly showed that the acquisition of both target items was a stochastically dependent process and not an independent one. That is, the values of the parameters that measured the acquisition of both the B and C response terms were not simply a function of the independent contribution of the parameters that measured the acquisition of each single response separately. Rather, the parameters that measured the acquisition of the entire associative cluster were well predicted by a stochastically dependent relationship between the single response parameters.

Obviously, these data rule out models in which memory representations consist of collections of independently associated concepts. Specifically, these results are inconsistent with multiple trace theories such as HAM and Ross and Bower's schema model, in which it is assumed that components of an associative cluster are represented in separate memory locations that are linked together in an independent fashion. These findings are, however, consistent with the unitary trace assumption that cues and targets are represented in a single holistic memory unit.

Although alternative multiple-trace interpretations are possible, it turns out that none of these explanations provide a particularly compelling or satisfactory account of the data. It could be argued, for example, that the dependence observed between the acquisition of both responses was simply an artifact of the preexperimental correlation between the A concept, the B concept, and the C concept. That is, it would be possible to observe spurious dependence between the acquisition of various concepts that are otherwise independently represented in memory, given that these concepts were already intercorrelated (i.e., had strong, well-learned, interconcept links) in memory. This explanation is tenuous at best. While it may provide a weak explanation of the data from conditions in which intracluster similarity was high, this reasoning leads to the prediction that as intracluster similarity decreases the probability of observing the independent acquisition of individual response terms should increase. That is, as intratriad typicality decreases (i.e., TT → AT → TA → AA → UR), the tendency to reject the null hypothesis of dependence should increase and the tendency to reject the null hypothesis of independence should decrease. As neither of these patterns were observed, this hypothesis is fundamentally inconsistent with the data.

Another explanation of these results, one that is consistent with the multiple trace claim that components of a cluster are independently represented in memory, is that the likelihoods of within-cluster links are highly intercorrelated. This argument has frequently been cited in the context of sentence-memory research. For example, in J. Anderson's ACT theory (1976, pp. 410-411), an "extreme model of individual variation" is proposed in which a strong committment to the independence assumption is abandoned. Here, differences between subjects interacts with differences between items such that some associations are formed with probability one while the remaining associations are formed with probability zero. It has already been pointed out that, " ... the appeal to individual differences is unmotivated ... (and, moreover) the individual differences argument commits ACT to an unparismonious treatment of sentence memory" (Goetz et al., 1981, p. 382). In general, such post-hoc modifications of the multiple trace theory amount to an extremely complicated version of the unitary trace position. More importantly, however, the weight of the evidence converges on the conclusion that the principles of relational association that operate to form unitary memory structures for item pairs are also operative in the formation of unitary memory structures for higher-order groupings.

Acquisition processes: Strength versus all-or-none.

The findings from both experiments were also quite clear with respect to the question of how acquisition processes operate during associative learning. Consistent with previous research in the paired-associate literature (e.g., Brainerd et al., 1980, 1981, 1982; Greeno, 1970; Greeno et al., 1978; Halff, 1977; Humphreys & Greeno, 1970), the data from both associative cluster experiments were in good statistical agreement with the three-state Markov model. Moreover, the degree of correspondence between the model and the data did not fluctuate as a function of whether items within a cluster were from the same semantic category (Experiment 2) or from different semantic categories (Experiment 1). In theoretical terms, this means that the acquisition of both related and unrelated associative triplets consisted of exactly two discrete stages. These data are important because they rule out the possible multiple trace interpretation that learning consists of a single continuous process. In particular, these results are inconsistent with the multiple trace prediction that goodness-of-fit should vary as a direct consequence of variation in the degree or strength of within-cluster similarity. Rather, these findings are consistent with the unitary trace assumption that learning consists of exactly three discrete performance states and that transitions between these states occur in an all-or-none manner.

As mentioned earlier, an explicit stage analysis of the acquisition process is absent in recent versions of the multiple trace theory. This omission may be due, in part, to the fact that much of recent research has relied on single-trial learning designs. When learning-to-criteiron designs are employed, the evidence overwhelmingly indicates that associative learning consists of a two-stage sequence. In order to account for this invariant structure in the learning data, post-hoc modifications must be incorporated into the multiple trace position. Because these models do contain implicit assumptions about the logical and tmeporal course of processes operating during acquisition, such changes are not entirely unwarranted, and indeed certain stage-like extrapolations would not be inconsistent with this position.

As suggested earlier, the most plausible two-stage multiple-trace interpretation can be found in the schema model. Here, learning consists of (1) storing or activating the concept nodes in memory and constructing or discovering a mediating schema, and (2) gradually incrementing the strength of trace-to-schema and schema-to-trace links. Consistent with the strength assumption in this model, it is predicted that the probability of a correct response should vary in a continous fashion from zero to one. This hypothesis, while inconsistent with the observation that transitions between states occurred in a discrete manner, could be salvaged by assuming that interstate transitions were actually mediated by a continuous buildup of response strength. That is, the discrete changes in the probability of a correct response (zero to $\underline{p}$, zero to one, and $\underline{p}$ to one) that were observed when the process made state-to-state transitions (U to P, U to T, and P to T, respectively), could have been the result of the operation of some kind of threshold process. This suggestion is not inconsistent with the two-stage hypothesis. For example, it is entirely possible to find that the three-state model provides a good overall account of the data but the stages themselves do not have the Markovian property (see Brainerd et al., 1982). That is, two-stage models do exist (e.g., M. Norman, 1964) where escape from the first-stage is mediated by a continuous process but second-stage escape is all-or-none, or first-stage escape is all-or-none and second-stage escape is mediated by a continuous process.

In order to investigate this theoretical claim, fine-grained aspects of the data were examined. The strength-threshold hypothesis implies that the observed probability of a correct response, rather than remaining constant across trials within a stage, should vary continuously until a discrete-change threshold is attained. Because the probability of a correct response is by definition zero in State U and by definition one in State T, this hypothesis was tested by examining variation in response probability in State P. When the distributions of two second-stage performance statistics, length of error runs and length of success runs, were examined, tests revealed that the probability of a correct response remained stationary across trials. More importantly, within-state stationarity did not fluctuate as a function of variation in the degree of interitem relatedness. That is, only discrete changes in the probability of a correct response were observed regardless of whether subjects were learning semantically unrelated lists (Experiment 1) or lists in which the associative clusters varied in their degree of category membership

(Experiment 2). These findings provide strong evidence against "theories that assume that improvements in recall are caused by gradual strengthening" (Brainerd et al., 1982, p. 638). In particular, the structure of these data not only rule out the strength assumption in the schema version of the multiple trace theory, but, more importantly, they rule out any post-hoc attempt to explain acquisition by invoking the concept of a threshold mechanism. In order to account for this pattern of results, such models would have to abandon the strength assumption. This rather unappealing alternative is made even more unattractive by the observation that such post-hoc deletions would result in a rather unwieldy version of the unitary trace position.

Loci of preexperimental organization effects.

The stages-of-learning analysis has revealed some rather interesting, and hitherto unreported, effects due to the manipulation of preexperimental knowledge. First, the manipulation of typicality in the absence of category membership (Experiment 1) did not have a substantial effect on the acquisition of associative clusters. While this is not particularly surprising, these results indicate that the influence of factors that may be incidentally correlated with degree of category membership (e.g., verbal attributes such as word length, frequency, or pronounceability) was negligible at best.

Second, the effects of manipulating the absence versus the presence of preexperimental knowledge (Experiments 1 versus 2) depended on whether learning or performance was being measured. With regard to learning, the results revealed that both first-stage (a' and a) and second-stage (b', b and d) learning parameters were affected.

If, as previous research has indicated, the first-stage learning parameters measure the difficulty of encoding and storage and the second-stage learning parameters measure the difficulty of learning to retrieve, then these data show that both the storage and retrieval components of acquisition are facilitated by the addition of preexperimental semantic knowledge. While these results were anticipated by most models of associative memory, they are inconsistent with the HAM version of the multiple trace theory. In particular, they are inconsistent with the claim that storage simply consists of activating concepts in isolation. Rather, these data support the contention that relations between concepts are important throughout the acquisition process, both at storage and at retrieval.

With regard to performance, on the other hand, the results indicated that even though performance was slightly better on the first trial in the intermediate state (as measured by the parameter 1-_e_) for related clusters, the majority of second-stage performance (as measured by the parameters _g_ and _h_) was no better for related than unrelated associative clusters. That performance remained essentially invariant regardless of whether items were related or unrelated is not inconsistent with any of the proposals considered in this paper. However, only one model, namely, the modified storage-retrieval version of the unitary trace theory, anticipated these results. According to this model, second-stage performance parameters measure an intermediate level of heuristic recall (i.e., context-free operations), whereas second-stage learning parameters measure the ease of acquiring item-specific retrieval algorithms (i.e., context-sensitive operations). Because preexperimental knowledge serves to increase item-specific information about members of each cluster, the addition of category membership should selectively affect algorithmic, but not heuristic, retrieval. Consistent with this interpretation, the performance parameters $\underline{g}$ and $\underline{h}$ did not react to the manipulation of preexperimental knowledge, however, the learning parameters $\underline{b}'$, $\underline{b}$, and $\underline{d}$ were influenced by whether items within a cluster were related or unrelated. It is not intuitively obvious how competing models of associative memory could account for the structure of these data. Given the compelling nature of these finding, as well as the lack of alternative explanations, the modified storage-retrieval model provides the most consistent account of the effects of preexperimental knowledge on acquisition.

Third, the effects of manipulating the degree of intracluster similarity (Experiment 2) also depended on whether learning or performance was being measured. In addition, the manipulation of typicality on the cue versus the target component of the associative clusters interacted with the learning parameters of the two stages. Contrary to the HAM version of the multiple trace theory, degree of category membership affected acquisition. That is, the results were consistent with the global hypothesis that as intracluster similarity increased, TT → AT → TA → AA, so too did ease of acquisition. Notice, however, that these effects were again confined to both first-stage and second-stage learning parameters, and only one second-stage learning parameter ($1-\underline{e}$). These results indicate that the effect of manipulating overall intracluster similarity was on learning rate and not on performance per se.

Next, consider the parameter-specific effects of manipulating typicality on the cue versus the target side of the associative triads. Here, three interesting patterns emerged. In terms of first-stage learning, analyses revealed that (1) the advantage of typical cues over atypical cues was confined to the first-trial learning parameter $\underline{a}'$, and (2) the advantage of typical targets over atypical targets was on both learning parameters $\underline{a}'$ and $\underline{a}$. That degree of cue and target category membership affected information storage is consistent with both the schema version of the multiple trace theory and the unitary trace models. A more interesting aspect of this result concerns the difference in the magnitude of this effect depending on whether typicality was manipulated on the cue or the target side of the triad. If, as Brainerd et al. (1981, p. 13) suggest, "prestorage difficulty should have more effect on early trials (Parameter $\underline{a}'$) than on later trials (Parameter $\underline{a}$)", then these results indicate that both cue and target typicality affected prestorage factors (i.e., encoding difficulty, but only target typicality affected the entire storage process.

Turning to the second-stage learning parameters, analyses revealed that (1) the parameter $\underline{b}'$ was negatively affected in the condition where both the cue and the targets were typical (TT) compared to the mixed cue and target conditions (AT and TA), and (2) the parameters $\underline{b}$ and $\underline{d}$ were positively affected in the mixed cue and target conditions (AT and TA) compared to the condition where both the cue and targets were atypical (AA). That both cue and target typicality affected second-stage retrieval learning is consistent with the schema model and the modified storage-retrieval model. However, these results are inconsistent with the cue-dependent interpretation of retrieval in the encoding-specificity and early storage-retrieval version of the unitary trace theory. The most fascinating aspect of these results, however, concerns the finding of what shall be called a retrieval-suppression effect. That is, compared to conditions in which the cues and targets were atypical, second-stage retrieval learning was facilitated when either the cue or the target component of a triad was typical (Effect 2). Such facilitation was not unexpected, and indeed, was predicted by two of the models discussed in this paper. Unexpectedly, however, second-stage retrieval learning was suppressed when both the cue and the targets were typical (Effect 1). That is, the parameter <u>b</u>' was significantly lower in the TT condition than in the AT or TA conditions. Recall that the parameter <u>b</u>' measures the probability of skipping the second-stage given the process escapes the first-stage of learning on Trial 1 (as measured by the parameter <u>a</u>'). That <u>a</u>' was higher in the TT condition than in the mixed list condition indicates that it was easier to encode and store a trace on Trial 1 when intracluster similarity was high. That <u>b</u>' was lower in the TT condition than in the mixed list conditions indicates that it was harder to retrieve information from a trace when intracluster similarity was high. Therefore, while it is easier to store highly similar items in memory, it is slightly more difficult to retrieve retrieve them.

It is not entirely clear how a multiple trace position would be compatible with these results. Specifically, it is difficult to imagine how any theory that posits that items are stored in separate memory locations and that associative retrieval routes are established in an independent fashion can account for the fact that interitem similarity facilitated retrieval when only ~~one or~~ two components of a cluster were typical category members, but inhibited ~~retrieval~~ when all three cluster components were typical category members. Alternatively, while this effect was not predicted by the unitary trace model, it is not difficult to see how this result can easily be incorporated in this theory. Recall that in the modified storage-retrieval version of the unitary trace theory, second-stage learning consists of developing an item-specific retrieval algorithm (Halff, 1977). This algorithm consists of two parts, a search component that locates the trace in memory and a decoding component that discriminates the target items for retrieval (Brainerd et al., 1981). It could be argued that while within-cluster similarity facilitates locating the trace in memory, it ~~also~~ inhibits within-trace item discrimination.

Such an interpretation is consistent with the interaction obtained here. That is, when within-trace similarity was high, as in the TT list condition, locating the trace in memory (i.e., search) was relatively easy but target discriminability (i.e., decoding) decreased, increasing the difficulty of retrieval. As the level of similarity decreased to an intermediate level, as in the mixed AT and TA list conditions, there would still be sufficient feature overlap to make search effective, however, target discriminability would increase thereby facilitating retrieval. When within-cluster similarity was low, as in the AA list condition, target discriminability would be relatively easy but due to the lower overall level of feature overlap, search would become more difficult, increasing retrieval difficulty. While it is obvious that this hypothethical interplay between the search and decoding components of algorithmic retrieval provides an adequate account of the second-stage learning data, it is also clear that this explanation is post-hoc. The important point to bear in mind, however, is that an explanatory mechanism already exists within the modified storage-retrieval framework, whereas one is absent in the multiple trace theory.

Turning to the second-stage performance parameters, the results once again indicated that even though performance was slightly better on the first trial in the second stage for clusters in which interitem similarity was high, the majority of second-stage performance was no better in high similarity conditions than in low similarity conditions. This result provides strong support for the modified storage-retrieval claim that second-stage performance is mediated by context-free heuristic retrieval operations.

Finally, there are three interesting parameter invariances that were observed regardless of the list being memorized. First, the parameter $\underline{d}$ was always greater than zero and the parameter $\underline{c}$ was always exactly zero. This result is inconsistent with the multiple trace claim that the probability of learning an item increases as a function of successfully retrieving that item. Rather, these data support the unitary trace contention that the probability of learning increases following an error. Consistent with this finding, a second relationship, namely, $\underline{g} \geq \underline{h}$, was also obtained in both experiments. That is, the probability of a success following an error was either greater than or equal to the probability of a success following a success in the second stage of learning. This relationship is also consistent with the unitary trace assumption that, "subjects rely on errors to tell them if further work needs to be done on an item that has escaped U" (Brainerd et al., 1981, p. 13).

Third, as in previous research, the relationship {$\underline{a}$', $\underline{a}$} $\leq$ {$\underline{b}$', $\underline{b}$, $\underline{d}$} was observed. What this relationship means is that first-stage learning was more difficult than second-stage learning. Theoretically, this relationship is fundamentally inconsistent with theories that claim that information storage is a relatively brief or insignificant aspect of associative learning (e.g., HAM). Rather, these data indicate that encoding and storage processes play a fundamental role in acquisition, one that is at least as important as association and retrieval.

Conclusion

Although most theorists agree that memory concepts consist of feature-bundles, there has been much debate concerning the appropriate structural format for representing this featural information in associative memory. Numerous theoretical models have been proposed, each of which can be classified according to whether concepts are represented in individual memory traces that are independently linked together (the multiple trace approach) or are represented in a single integral memory trace (the unitary trace approach). While it has typically proved difficult to obtain unambiguous evidence that favors one approach over the other, the results of the present study provide convergent evidence that associative memory consists of holistic memory units rather than independent memory structures.

Similarly, while most theorists would agree that associative learning consists of encoding, storage and retrieval, there is little agreement on whether these processes operate in a continuous or discrete manner. Unfortunately, many of the models that have been proposed have failed to provide an adequate mathematical account of the structure of learning data. The analyses that were used in the present research unequivocally show that when subjects memorize lists of words to some stringent acquisition criterion, the data are three-state Markovian in nature. That is, learning consists of exactly two stages, performance within stages is stationary rather than nonstationary, and learning (i.e., transitions between stages) occurs in an all-or-none discrete manner rather than in a continuous fashion.

Finally, while the use of a stages-of-learning approach does not entail a committment to any particular theoretical orientation, it does provide an analytically precise framework within which various alternatives can be tested. Specifically, because the available evidence suggests that, "the first-stage and second-stage transition parameters (react) appropriately to manipulations designed to affect storage and retrieval respectively" (Halff, 1977, pp. 385-386), tests of different theoretical interpretations of the effects of experimental manipulations on encoding, storage and retrieval can be conducted. In addition, it can be determined whether manipulations affect learning, performance, or both. It was because of the sensitivity of these analyses that the precise locus of the effects of preexperimental knowlege on learning could be obtained. Indeed, it was only when these fine-grained parameter analyses were examined that the various versions of the multiple and unitary trace interpretations could be distinguished empirically. Consistent with the more global analyses that were conducted, detailed investigation of the pattern of parameterwise results converged on a single theroetical interpretation, namely, the modified storage-retrieval version of the unitary trace model.

REFERENCES

Anderson, J. R. Language, memory and thought. Hillsdale, N.J.: Erlbaum, 1976.

Anderson, J. R. and Bower, G. H. On an associative trace for sentence memory. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 673-680.

Anderson, J. R. and Bower, G. H. Configural properties in sentence memory. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 594-605.

Anderson, J. R. and Bower. G. H. Human associative memory. Washington, D.C.: Winston, 1973.

Anderson, R. C. and Ortony, A. On putting apples into bottles: A problem of polysemy. Cognitive Psychology, 1975, 7, 167-180.

Anderson, T. W. and Goodman, L. A. Statistical inference about Markov chains. Annals of Mathematical Statistics, 1957, 28, 89-110.

Asch, S. E. A reformulation of the problem of association. American Psychologist, 1969, 24, 92-102.

Battig, W. F. and Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology Monograph, 1969, 80(No. 3, Pt. 2).

Bernbach, H. A. Derivation of learning process statistics for a general Markov model. Psychometrika, 1966, 31, 225-243.

Bousfield, W. A. The occurrence of clustering in recall of randomly arranged associates. Journal of General Psychology, 1953, 49, 229-240.

Bower, G. H. A multicomponent theory of the memory trace. In K. W. Spence and J. T. Spence (Eds.), The psychology of learning and motivation (Vol. 1). New York: Academic Press, 1967.

Bower, G. H. and Theios, J. A learning model for discrete performance levels. In R. C. Atkinson (Ed.), Studies in mathematical psychology. Stanford, California: Stanford University Press, 1964.

Brainerd, C. J., Desrochers, A., and Howe, M. L. Stages-of-learning analysis of picture-word effects in associative memory. Journal of Experimental Psychology: Human Learning and Memory, 1981, 7, 1-14.

Brainerd, C. J. and Howe, M. L. Stages-of-learning analysis of developmental interaction in memory, with illustrations from interactions in picture-word effects. Developmental Review, 1982, 2, in press.

Brainerd, C. J., Howe, M. L. and Desrochers, A. Interpreting associative-learning stages. Journal of Experimental Psychology: Human Learning and Memory, 1980, 6, 754-765.

Brainerd, C. J., Howe, M. L. and Desrochers, A. The general theory of two-stage learning: A mathematical review with illustrations from memory development. Psychological Bulletin, 1982, 91, 634-665.

Bransford, J. D. and Franks, J. J. The abstraction of linguistic ideas. Cognitive Psychology, 1971, 2, 331-350.

Burke, C. J. and Rosenblatt, M. A Markovian function of a Markov chain. Annals of Mathematical Statistics, 1958, 29, 1112-1122.

Chernoff, H. and Lehmann, E. L. Use of maximum likelihood test in chi square tests of goodness of fit. Annals of Mathematical Statistics, 1954, 25, 579-586.

Cofer, C. N. On some factors in the organizational characteristics of free recall. American Psychologist, 1965, 20, 261-272.

Cofer, C. N. Associative overlap and category membership as variables in paired associate learning. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 230-235.

Craik, F. I. M. and Lockhart, R. S. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 1971, 11, 671-684.

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), Psychology: A study of a science (Vol. 2). New York: McGraw-Hill, 1959.

Estes, W. K. The information-processing approach to cognition: A confluence of metaphors and methods. In W. K. Estes (Ed.), Handbook of learning and cognitive processes (Vol. 5). Hillsdale, N.J.: Erlbaum, 1978.

Estes, W. K. and DaPolito, F. Independent variation of information storage and retrieval processes in paired-associate learning. Journal of Experimental Psychology, 1967, 75, 18-26.

Fillmore, C. J. The case for case. In E. Bach and R. T. Harms (Eds.), Universals in linguistic theory. New York: Holt, Rinehart and Winston, 1968.

Flexser, A. J. and Tulving, E. Retrieval independence in recognition and recall. Psychological Review, 1978, 85, 153-171.

Foss, D. J. and Harwood, D. A. Memory for sentences: Implications for human associative memory. Journal of Verbal Learning and Verbal Behavior, 1975, 14, 1-16.

Franks, J. J. and Bransford, J. D. The acquisition of abstract ideas. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 311-315.

Goetz, E. T., Anderson, R. C. and Schallert, D. L. The representation of sentences in memory. Journal of Verbal Learning and Verbal Behavior, 1981, 20, 369-385.

Greeno, J. G. Paired-associate learning with short-term retention: Mathematical analysis and data regarding identification of parameters. Journal of Mathematical Psychology, 1967, 4, 430-472.

Greeno, J. G. Identifiability and statistical properties of two-stage learning with no successes in the initial stage. Psychometrika, 1968, 33, 173-216.

Greeno, J. G. How associations are memorized. In D. A. Norman (Ed.), Models of human memory. New York: Academic Press, 1970.

Greeno, J. G., James, C. T., DaPolito, F. and Polson, P. G. Associative learning: A cognitive analysis. Englewood Cliffs, N.J.: Prentice-Hall, 1978.

Greeno, J. G. and Steiner, T. E. Markovian processes with identifiable states: General considerations and applciations to all-or-none learning. Psychometrika, 1964, 29, 309-333.

Halff, H. M. The role of opportunities for recall in learning to retrieve. American Journal of Psychology, 1977, 90, 383-406.

Hintzman, D. L. Simpson's paradox and the analysis of memory retrieval. Psychological Review, 1980, 87, 398-410.

Humphreys, M. S. and Greeno, J. G. Interpretation of paired-associate memorizing. Journal of Mathematical Psychology, 1970, 7, 275-292.

Jenkins, J. J., Mink, W. D. and Russell, W. D. Associative clustering as a function of verbal associative strength. Psychological Reports, 1958, 4, 127-136.

Jenkins, J. J. and Russell, W. A. Associative clustering during recall. Journal of Abnormal and Social Psychology, 1952, 47, 818-821.

Jones, G. V. A fragmentation hypothesis of memory: Cued recall of pictures and of sequential position. Journal of Experimental Psychology: General, 1976, 105, 277-293.

Jones, G. V. Repeated cueing and the structure of recall. British Journal of Mathematical and Statistical Psychology, 1978, 31, 1-10. (a)

Jones, G. V. Tests of a structural theory of the memory trace. British Journal of Psychology, 1978, 69, 351-368. (b)

Kamman, R. Associability: A study of the properties of associative ratings and the role of associations in word-word learning. Journal of Experimental Psychology Monograph, 1968, 78, (No. 4, Pt. 2).

Kemeny, J. G. and Snell, J. L. Finite Markov chains. Princeton, N.J.: Van Nostrand, 1960.

Kintsch, W. All-or-none learning and the role of repetition in paired-associate learning. Science, 1963, 140, 310-312.

Kintsch, W. The representation of meaning in memory. Hillsdale, N.J.: Erlbaum, 1974.

Kintsch, W. and Morris, C. J. Application of a Markov model to free recall and recognition,. Journal of Experimental Psychology, 1965, 69, 200-206.

Koffka, K. Principles of gestalt psychology. New York: Harcourt, Brace & World, 1935.

Kohler, W. Dynamics in psychology. New York: Liverwright, 1940.

Kohler, W. On the nature of associations. Proceedings of the American Philosophical Society, 1941, 84, 489-502.

Kohler, W. Gestalt psychology. New York: Liverwright, 1947, authorized paperback reprint, New York: New American Library.

Levine, G. and Burke, C. J. Mathematical model techniques for learning theories. New York: Academic Press, 1972.

Mandler, G. Organization and memory. In K. W. Spence and J. T. Spence (Eds.), The psychology of learning and motivation (Vol. 1). New York: Academic Press, 1967.

Marschark, M. and Paivio, A. Integrative processing of concrete and abstract sentences. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 217-231.

Marshall, G. R. Stimulus characteristics contributing to organization in free recall. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 364-374.

Martin, E. M. Simpson's paradox resolved: A reply to Hintzman. Psychological Review, 1981, 88, 372-374.

Mervis, C. B., Catlin, J. and Rosch, E. Relationships among goodness-of-example, category norms, and word frequency. Bulletin of the Psychonomic Society, 1976, 7, 283-284.

Millward, R. B. Derivations of learning statistics from absorbing Markov chains. Psychometrika, 1969, 34, 215-232.

Minsky, M. A framework for representing knowledge. In P. H. Winston (Ed.), The psychology of computer vision. New York: McGraw-Hill, 1975.

Norman, D. A. and Rumelhart, D. E. Explorations in cognition. San Francisco: Freeman, 1975.

Norman, M. F. A two-phase model and an application to verbal discrimination learning. In R. C. Atkinson (Ed.), Studies in mathematical psychology. Stanford, Calif.: Stanford University Press, 1964.

Pagel, J. C. A Markov analysis of transfer in paired-associate learning with high intralist similarity. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 456-470.

Pelligrino, J. W. and Salzberg, P. M. Encoding specificity in cued recall and context recognition. Journal of Experimental Psychology: Human Learning and Memory, 1975, 1, 261-270.

Polson, P. G. Statistical methods for a general theory of all-or-none learning. Psychometrika, 1970, 35, 51-72.

Postman, L. The effects of language habits on the acquisition and retention of verbal associations. Journal of Experimental Psychology, 1962, 64, 7-19.

Postman, L., Frazer, J. and Burns, G. Unit-sequence facilitation in recall. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 217-224.

Restle, F. and Greeno, J. G. Introduction to mathematical psychology. Reading, Mass.: Addison-Wesley, 1970.

Rosch, E. Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 1975, 104, 192-233.

Rosch, E. and Mervis, C. B. Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 1975, 7, 573-605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. and Boyes-Braem, P. Basic objects in natural categories. Cognitive Psychology, 1976, 8, 382-439.

Rosch, E., Simpson, C. and Miller, R. S. Structural bases of typicaltiy effects. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 491-502.

Ross, B. H. and Bower, G. H. Comparisons of models of associative recall. Memory and Cognition, 1981, 9, 1-16.

Rumelhart, D. E., Lindsay, P. N. and Norman, D. A. A process model for long-term memory. In E. Tulving and W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Schacter, D. L., Eich, J. E. and Tulving, E. Richard Semon's theory of memory. Journal of Verbal Learning and Verbal Behavior, 1978, 17, 721-743.

Schank, R. and Abelson, R. Scripts, plans, goals, and understanding. Hillsdale, N.J.: Erlbaum, 1977.

Siddal, J. N. and Bonham, D. J. OPTISEP: Optimization subroutine package. Department of Mechanical Engineering, McMaster University, 1974.

Theios, J. Finite integer models for learning in individual subjects. Psychological Review, 1968, 75, 292-307.

Theios, J., Leonard, D. W. and Brelsford, J. W., Jr. Hierarchies of learning models that permit likelihood ratio comparisons. Journal of Experimental Psychology: General, 1977, 106, 213-225.

Toglia, M. P. and Battig, W. F. Handbook of semantic word norms. Hillsdale, N.J.: Erlbaum, 1978.

Tulving, E. Relation between encoding specificity and levels of processing. In L. S. Cermak and F. I. M. Craik (Eds.), Levels of processing in human memory. Hillsdale, N.J.: Erlbaum, 1979.

Underwood, B. J. Attributes of memory. Psychological Review, 1969, 76, 559-573.

Underwood, B. J., Runquist, W. N. and Schulz, R. W. Response learning in paired-associate lists as a function of intralist similarity. Journal of Experimental Psychology, 1959, 58, 70-78.

Underwood, B. J. and Schulz, R. W. Meaningfulness and verbal learning. Philadelphia: Lippincott, 1960.

FOOTNOTES

[1] While much has been made of the flexibility of labeled over unlabeled associations, in most list-learning situations labels are omitted and connections between nodes which consist of several links are summarized by just one (e.g., Ross & Bower, 1981). In fact, it has been shown that labeled associations are unnecessary as they can always be rewritten as unlabeled graphs simply by defining the label itself as a node (see Rumelhart, Lindsay & D. Norman, 1972). In essence, the addition of labels to links turns out to be a matter of computational efficiency (i.e., a labeled associative network can be searched more efficiently than an unlabeled one) rather than an empirical necessity (see J. Anderson & Bower, 1973). In a similar vein, abstract nodes tend to be considered isomorphic to their verbal labels in discussion of verbal learning experiments (e.g., J. Anderson & Bower, 1973, Chapter 14). It is not clear, therefore, what the pragmatic advantages are of replacing words with nodes.

[2] It is not clear whether this applies exclusively to the computer simulation of learning (as it is apparent that list items, if they are words, sentences, etc., are already represented in the system and are automatically activated upon presentation), or whether it is also meant to extend to empirical contexts. For the sake of consistency, it will be assumed that the simulation and experimental conditions are meant to be analogous, and that as long as the items presented are "common" words (the distinction made by Anderson and Bower is one between words as opposed to nonword CVCs and CCCs) subjects in an experiment will automatically

access and activate (i.e., parse and match) mental representations of the list items in memory, obviating the need to learn them.

3 Actually, these tests as well as subsequent likelihood ratio tests, were computed using an equivalent statistical procedure. Because the estimation of the likelihood of the data involved the minimization of twice the negative natural log of the likelihood function, these ratios simply reduce to the subtraction of the likelihood with the greater number of degrees of freedom from the likelihood with the smaller number of degrees of freedom (see Brainerd et al., 1982). In the present case, the likelihood ratio test simply reduces the subtraction of $-2\log_e L_8$ from $-2\log_e L_5$, or $2\log_e L_5 - 2\log_e L_8$.

APPENDIX A

The Three-State Markov Model of Two-Stage Learning

The general model that was used in the analysis of the data is a three-state absorbing Markov chain. The notation used to describe this model closely follows that of Greeno (1968) and Brainerd et al. (1982).

The four major assumptions about learning that are associated with this model are: (1) learning involves exactly two stages; (2) there are three discrete performance levels or states, U, P, and T, in which a correct response occurs with probabilities zero, $0 < p < 1$, and one, respectively; (3) transitions between these states have the Markov property (i.e., they occur on at most one trial or are all-or-none); and (4) there are no backward transitions such that once the process has escaped a particular state it cannot re-enter it. This class of models has two preabsorbing states, an initial "unlearned" State U in which errors occur with probability one, and an intermediate "partially learned" State P in which both errors and successes occur with average probabilities $1-p$ and p, respectively. The third state is a terminal, absorbing "learned" State T in which successes occur with probability one.

Kemeny and Snell (1960) have shown that all the information in such models can be conveyed by an elementary matrix equation that consists of a row vector, a square matrix, and a column vector. The canonical form of the model has the following starting vector, transition matrix, and correct response vector:

$$R = \{T(1),\ P_E(1),\ P_C(1),\ U(1)\} = \{\underline{t},\ (1-\underline{s}-\underline{t})\underline{r},\ (1-\underline{s}-\underline{t})(1-\underline{r}),\ \underline{s}\};$$

$$\underline{M} = \begin{array}{c} \\ T(\underline{n}) \\ P_E(\underline{n}) \\ P_C(\underline{n}) \\ U(\underline{n}) \end{array} \begin{array}{c} \begin{array}{cccc} T(\underline{n}+1) & P_E(\underline{n}+1) & P_C(\underline{n}+1) & U(\underline{n}+1) \end{array} \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ \underline{d} & (1-\underline{d})(1-\underline{g}) & (1-\underline{d})\underline{g} & 0 \\ \underline{c} & (1-\underline{c})(1-\underline{h}) & (1-\underline{c})\underline{h} & 0 \\ \underline{ab} & \underline{a}(1-\underline{b})\underline{e} & \underline{a}(1-\underline{b})(1-\underline{e}) & 1-\underline{a} \end{array}\right] \end{array} ;C = \left[\begin{array}{c} 1 \\ 0 \\ 1 \\ 0 \end{array}\right], \quad (A1)$$

where P_E and P_C represent error and success substates, respectively, in the intermediate State P. The row vector, or <u>starting vector</u> R, gives the probability of starting in each of the three possible states. The square matrix, or <u>transition matrix</u> M, gives the probability of being in any state or Trial <u>n</u>+1 conditional on being in any other state on Trial <u>n</u>. Finally, the column vector, or <u>response vector</u> C, gives the probability of a correct response in each of the three possible states.

In this Markovian representation of learning, acquisition is viewed in terms of three possible sequences of state changes. First, subjects may start in State U and learn by simply making an all-or-none transition from U to T. Second, subjects may start in State U and learn by first making an all-or-none transition from U to P and then making an all-or-none transition from P to T. Third, subjects may start in State P and learn by making an all-or-none transition from P to T.

These various acquisisition sequences can be described in terms of the theoretical parameters in Equation A1 (for a more detailed account, see Brainerd et al., 1982). First, consider an acquisition sequence that begins in the first state. The probability that the learning process escapes U on the first learning trial is $\underline{a}'$ (where $\underline{a}' = 1-\underline{s}$). On any learning trial after Trial 1 in which the process was in U at the start of the trial, it either escapes U with probability $\underline{a}$ or remains in U with probability $1-\underline{a}$. If the process escapes U on Trial 1, it either goes directly to the learned State T with probability $\underline{b}'$ [where $\underline{b}' = \underline{t}/(1-\underline{s})$] or it enters the partially learned State P with probability $1-\underline{b}'$. If the process escapes U on any trial after Trial 1, it either goes directly to T with probability $\underline{b}$ or it enters P with probability $1-\underline{b}$.

Next, consider an acquisition sequence that either begins in the partially learned State P or has made a U-to-P transition. In the former case, the first trial in P is an error trial P_E with probability $(1-\underline{s}-\underline{t})\underline{r}$ or is a success trial P_C with probability $(1-\underline{s}-\underline{t})(1-\underline{r})$. In the latter case, the first trial is a P_E trial with probability $\underline{e}$ or a P_C trial with probability $1-\underline{e}$. After each learning trial in substate P_E, the process either escapes to T with probability $\underline{d}$ or remains in P with probability $1-\underline{d}$. Given that the process remains in P, the next trial is a P_E trial with probability $1-\underline{g}$ or a P_C trial with probability $\underline{g}$. After each learning trial in substate P_C, the process either escapes to T with probability $\underline{c}$ or remains in P with probability $1-\underline{c}$. Given that the process remains in P, the next trial is a P_E trial with probability $1-\underline{h}$ or a P_C trial with probability $\underline{h}$. On any trial in P, whether it is a P_E or P_C trial, the probability of falling back to U is zero. Similarly, if the process begins in the learned State T (with probability $\underline{t}$) or transits there from either U or P, it remains there with probability one as the probability of falling back to either

U or P is zero (i.e., State T is an absorbing state).

Equation A1 permits examination of at least three possible loci of effects due to experimental manipulations. The effects of manipulations on the first stage of learning (State U) can be measured by examining changes in the <u>learning</u> parameters <u>a'</u> and <u>a</u>. The effects of manipulations on the second stage of learning (State P) can be measured in two ways. First, the effects of manipulations on second-stage learning can be measured by examining changes in the <u>learning</u> parameters <u>b'</u>, <u>b</u>, <u>c</u>, and <u>d</u>. Second, the effects of manipulations on second-stage performance can be measured by examining changes in the <u>performance</u> parameters <u>r</u>, <u>e</u>, <u>g</u>, and <u>h</u>.

The first step in the application of Equation A1 is to obtain numerical estimates of both the learning and performance parameters. Unfortunately, it turns out that the parameters of Equation A1 cannot be estimated directly from the data. This is because the states in the theoretical model in Equation A1 are only partially observable. Because more detailed discussions of this problem are readily available (e.g., Brainerd et al., 1982; Greeno, 1967, 1968; Levine & Burke, 1972, Chapter 8; Restle & Greeno, 1970, Chapter 10), only a summary of the important concepts will be presented here.

Briefly, there are various ways in which the events postulated in a model can make contact with the corresponding events in the experimental outcome space (in this case, as in most learning situations, the outcome space is simply strings of errors and successes). First, every unique event in the experimental outcome space may correspond to a unique event in the model. In this case, the model is said to have observable states because the mapping of the data onto the model is one-to-one. Second, every unique event in the experimental outcome space may correspond to more than one event in the model. In this case, the model is said to have nonobservable states because the mapping of the data onto the model is one-to-many. Finally, intermediate cases, like the one here, exist in which the mapping of the data onto the model is one-to-one for some aspects of the experimental outcome space, but one-to-many for other aspects. These models are said to have partially observable states because not all of the performance states (in this case U, P_E, P_C, and T) can be uniquely differentiated in the experimental outcome space. In terms of the three-state model in Equation A1, any error before the first success cannot be unambiguously classified as an instance of U because such errors, according to the model, may be an instance of either P_E or U. Similarly, any success after the last error or successes in protocols with no errors cannot be unambiguously classified as an instance of T because such successes, according to the model, may be an instance of either P_C or T. However, any error after the first success must, by definition, be an instance of P_E, and any success before the last error must, by definition, be an instance of P_C. Because, as shown above, U and T are nonobservable and P_E and P_C are partially observable, the model described in Equation A1 is a system that contains only partially observable states.

The importance of this distinction between observable, partially observable, and nonobservable states, lies in the fact that the parameters of an observable states model are identifiable. The advantage of identifiable parameters is that they can be directly estimated from the data. That is, because every response in an observable-states system can be unambiguously classified, the data from an experiment will yield a unique estimate of each parameter. On the other hand, although it is possible for models that are either nonobservable or partially observable to have identifiable parameters, these models also contain, by definition, parameters that are only partially identifiable or completely nonidentifiable. The problem with these latter parameters is that they cannot be directly estimated from the data of an experiment. In situations such as this, the property of identifiability must be investigated mathematically before the model can be applied to the data.

The construction of an identity proof has been greatly simplified for Markov models (see, Burke & Rosenblatt, 1958; Kemeny & Snell, 1960) so that it is possible to investigate the identifiability of a model whose states are not completely observable by constructing a new observable-states process that is implied by the model. By expressing the identifiable parameters of this new observable-states process as algebraic functions of the parameters of the original model, the question of the identifiability of the latter model's parameters can be determined. Since the states in Equation A1 are only partially observable, parameter identifiability must be investigated by constructing a system with observable states. The values of the parameters in Equation A1 can then be indirectly estimated by first estimating the values of the parameters of an observable states process that is implied by Equation A1, and then solving a system of simultaneous equations that map the theoretical parameters of Equation A1 onto the identifiable parameters of the observable-states model.

The starting vector, transition matrix, and correct response vector of the observable Markov chain implied by Equation A1 can be found in Equation A2 below.

$$\underline{R} = \{Q(1),\ R(1),\ S(1),\ I_1(1),\ I_2(1),\ \ldots,\ I_j(1)\} = \{\pi,\ 0,\ (1-\pi)(1-\theta),\ (1-\pi)\theta,\ 0,\ \ldots,\ 0\};$$

$$M = \begin{array}{r|ccccccc|}
 & Q(\underline{n}+1) & R(\underline{n}+1) & S(\underline{n}+1) & I_1(\underline{n}+1) & I_2(\underline{n}+1) & \ldots & I_j(\underline{n}+1) \\ \hline
Q(\underline{n}) & 1 & 0 & 0 & 0 & 0 & \ldots & 0 \\
R(\underline{n}) & \underline{u} & (1-\underline{u})\underline{v} & (1-\underline{u})(1-\underline{v}) & 0 & 0 & \ldots & 0 \\
S(\underline{n}) & 0 & \underline{z} & 1-\underline{z} & 0 & 0 & \ldots & 0 \\
I_1(\underline{n}) & \alpha_1 & 0 & \beta_1 & 0 & (1-\alpha_1-\beta_1) & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ldots & \vdots \\
I_{\underline{j}-1}(\underline{n}) & \alpha_{\underline{j}-1} & 0 & \beta_{\underline{j}-1} & 0 & 0 & \multicolumn{2}{c|}{(1-\alpha_{\underline{j}-1}-\beta_{\underline{j}-1})} \\ \hline
\end{array}\ ; C = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} . \qquad (A2)$$

The new observable states in Equation A2, as defined in Brainerd et al. (1982, p. 643), are as follows:

Q = the state on all trials in protocols with no errors and the state on all trials after the last error in protocols with one or more errors;

S = the state on all correct response trials before the last error in protocls with one or more errors;

R = the state on all error trials after the first correct response trial in protocls with one or more errors;

I_1 = the state on Trial 1 if an error occurs on that trial;

I_2 = the state on Trial 2 if an error occurs on that trial and if an error occured on Trial 1;

I_j = the state on Trial j if an error occurs on that trial and if an error occurred on each of the j-1 preceding trials, where j is the maximum length of the initial error run observed in the outcome space.

An advantage of an observable-states model is the availability of a likelihood function that expresses the a posteriori probability of any given set of data in terms of the model's parameters (Theios, 1968; Theios et al., 1977). Given the availability of such a function, algebraic expressions for the maximum-likelihood estimators of each identifiable parameter can usually be derived. This is done by taking the log of the likelihood function, differentiating, and minimizing the partial derivatives of the various parameters (T. Anderson and Goodman, 1957). The likelihood function for Equation A2 is,

$$L = (\pi)^{\underline{N}(Q)} (1-\pi)^{\underline{N}(S)+\underline{N}(I_1)} (\theta)^{\underline{N}(I_1)} (1-\theta)^{\underline{N}(S)} (\underline{u})^{\underline{N}(R,Q)} (1-\underline{u})^{\underline{N}(R,R)+\underline{N}(R,S)} (\underline{v})^{\underline{N}(R,R)} (1-\underline{v})^{\underline{N}(R,S)} (\underline{z})^{\underline{N}(S,R)} (1-\underline{z})^{\underline{N}(S,S)} \{\prod_{\underline{i}=1}^{j-1} (\alpha_{\underline{i}})^{\underline{N}(I_{\underline{i}},Q)} (\beta_{\underline{i}})^{\underline{N}(I_{\underline{i}},S)} (1-\alpha_{\underline{i}}-\beta_{\underline{i}})^{\underline{N}(I_{\underline{i}},I_{\underline{i}+1})}, \qquad \text{(A3)}$$

where $\underline{N}(\underline{m})$ is the number of times that the process is observed to start in state $\underline{m}$ ($\underline{m}$ = Q, S, I), $\underline{N}(\underline{m},\underline{n})$ is the number of times that the process is observed to go from state $\underline{m}$ on Trial $\underline{i}$ ($\underline{m}$ = Q, R, S, I_1, I_2, ..., I_{j-1}) to state $\underline{n}$ on Trial $\underline{i}$+1 ($\underline{n}$ = Q, R, S, I_2, I_3, ..., I_j).

The set of identifiable parameters for an observable-states model is usually called a parameterization of that model. While several parameterizations exist, the one that was used here was derived by Brainerd et al. (1980, 1981, 1982). The fact that Equations A1 and A2 are equivalent means that a set of functions can be written which map the 10 theoretical parameters of Equation A1 (s, t, r, a, b, c, d, e, g, h) onto the 8 identifiable parameters of Equation A2 (π, θ, u, v, z, w, ρ, γ). It is this parameterization that permits the equivalence of the data space of Equation A2 and the theoretical state space of Equation A1. It is this space, and not the states, that are mapped onto one another by the algebraic mapping functions. The definitions of the identifiable parameters in Equation A2, in terms of the theoretical parameters of Equation A1, are:

$$\pi = t + [(1-s-t)(1-r)c]/[1-(1-c)h]; \tag{A4}$$

$$\theta = [s + (1-s-t)r]/\langle 1-t-\{(1-s-t)(1-r)c/[1-(1-c)h]\}\rangle; \tag{A5}$$

$$u = d + [(1-d)cg]/[1-(1-c)h]; \tag{A6}$$

$$v = \{(1-g)[1-(1-c)h]\}/\{(1-g)[1-(1-c)h] + g(1-c)(1-h)\}; \tag{A7}$$

$$z = 1-(1-c)h\ ; \tag{A8}$$

$$w = 1-a\ ; \tag{A9}$$

$$\rho = \langle as\{b + (1-b)(1-e)[c/(1-(1-c)h]\} + (1-s-t)r[ch + (1-g)d]/[1-(1-c)h]\rangle/\langle as[1-(1-b)e] + (1-s-t)r[g + (1-g)d]\rangle; \tag{A10}$$

$$\gamma = \{s[1-a + a(1-b)e] + (1-s-t)r[1-d(1-g)]\}/[s + (1-s-t)r]. \tag{A11}$$

The first 5 identifiable parameters appear in the observable-states system in Equation A2. The 3 remaining parameters, w, ρ, and γ, are complex functions of all the α_i and β_i, which are defined as:

$$\alpha_i = u - \frac{(w)^{i-1}[u - (1-\gamma)\rho]}{\dfrac{(w)^{i-1}[\gamma-(1-u)v]}{w-(1-u)v} + \left[1 - \dfrac{\gamma-(1-u)v}{w-(1-u)v}\right]\left[(1-u)v\right]^{i-1}} \quad \text{and} \tag{A12}$$

$$\beta_{\underline{i}} = (1-\underline{u})(1-\underline{v}) - \frac{(\underline{w})^{\underline{i}-1}[\gamma-(1-\underline{u})\underline{v}]\{1-[\underline{u}-(1-\gamma)\rho]/[\gamma-(1-\underline{u})\underline{v}]\}}{\dfrac{(\underline{w})^{\underline{i}-1}[\gamma-(1-\underline{u})\underline{v}]}{\underline{w}-(1-\underline{u})\underline{v}} + \left[1 - \dfrac{\gamma-(1-\underline{u})\underline{v}}{\underline{w}-(1-\underline{u})\underline{v}}\right][(1-\underline{u})\underline{v}]^{\underline{i}-1}} \quad \text{(A13)}$$

The likelihood function for the 8 identifiable parameters can now be written by substituting the definitions in Equations A12 and A13 for the $\alpha_{\underline{i}}$ and $\beta_{\underline{i}}$ in Equation A3. When the log of the revised likelihood function is taken, and then differentiated, expressions can be derived for the partial derivatives of three parameters, namely, π, θ, and $\underline{z}$:

$$\frac{\partial \log L}{\partial \pi} = \frac{\underline{N}(Q)}{\pi} - \frac{\underline{N}(Q) + \underline{N}(S) + \underline{N}(I_1)}{1-\pi} \quad \text{(A14)}$$

$$\frac{\partial \log L}{\partial \theta} = \frac{\underline{N}(I_1)}{\theta} - \frac{\underline{N}(I_1) + \underline{N}(S)}{1-\theta} \quad \text{(A15)}$$

$$\frac{\partial \log L}{\partial \underline{z}} = \frac{\underline{N}(S,R)}{\underline{z}} - \frac{\underline{N}(S,S)}{1-\underline{z}} \quad \text{(A16)}$$

When the expressions on the right are minimized, the maximum-likelihood estimators of the three parameters are,

$$\hat{\pi} = \frac{\underline{N}(Q)}{\underline{N}(Q) + \underline{N}(S) + \underline{N}(I_1)} \quad \text{(A17)}$$

$$\hat{\theta} = \frac{\underline{N}(I_1)}{\underline{N}(I_1) + \underline{N}(S)} \quad \text{(A18)}$$

$$\hat{\underline{z}} = \frac{\underline{N}(S,R)}{\underline{N}(S,R) + \underline{N}(S,S)} \quad \text{(A19)}$$

Since algebraic expressions cannot be derived for the remaining five parameters, their maximum-likelihood estimates must be found by numerical methods. This was done using the OPTISEP (Siddal & Bonham, 1974) computerized search routine.

APPENDIX B

Theoretical Expressions for the Four Random Variables used to Evaluate the Sufficiency of the Three-State Model and the Two Random Variables used to Assess Second-Stage Performance

When examining the sufficiency of the two-stage model, it is advisable to study three statistical groups: (1) Statistics that focus on State U; (2) Statistics that focus on State P; and (3) Statistics that focus on the process as a whole. The following expressions for the sampling distributions of these statistics, as well as those used to evaluate second-stage performance, are presented below in terms of the 8 identifiable parameters of Equation A2. The matrix algorithm techniques used in the derivation of these statistics can be found in Bernbach (1966), Millward (1969) and Levine and Burke (1972) (see also Brainerd et al., 1982; Greeno, 1968).

State U

Initial error run. Since, by definition, only errors are possible in State U, the appropriate statistic is the distribution of errors before the first success. Let $\underline{C}$ be a random variable that counts the number of errors before the first correct response. The sampling distribution of $\underline{C}$ is given by,

$$P(\underline{C} = \underline{k}) = \begin{cases} \pi + (1-\pi)(1-\theta) & \text{for } \underline{k} = 0, \text{ and} \\ (1-\pi)\theta \dfrac{[\gamma-(1-\underline{u})\underline{v}](1-\underline{w})\underline{w}^{\underline{k}-1}}{\underline{w}-(1-\underline{u})\underline{v}} + \left[1 - \dfrac{\gamma-(1-\underline{u})\underline{v}}{\underline{w}-(1-\underline{u})\underline{v}}\right][1-(1-\underline{u})\underline{v}] & \\ [(1-\underline{u})\underline{v}]^{\underline{k}-1} & \text{for } \underline{k} = 1, 2, 3, \ldots \end{cases} \tag{B1}$$

State P

For State P, two staistics are given, length of error runs after the first success and length of success runs before the last error, that measure the stability of stage-two performance, and a third statistic, total number of errors after first success, that is related to stage-two learning.

Length of error runs after the first success. Let LR be a random variable that counts the number of consecutive errors in any run of errors after the first correct response has occurred. The distribution of LR is given by,

$$P(\underline{LR} = \underline{k}) = \begin{cases} \underline{u} + (1-\underline{u})(1-\underline{v}) & \text{for } \underline{k} = 1, \text{ and} \\ [(1-\underline{u})\underline{v}]^{\underline{k}-1}[\underline{u} + (1-\underline{u})(1-\underline{v})] & \text{for } \underline{k} = 2, 3, \ldots . \end{cases} \tag{B2}$$

Length of success runs before the last error. Let LS be a random variable that counts the number of consecutive errors in any run of correct responses that occurs before the last error. The distribution of LS is given by,

$$P(\underline{LS} = \underline{k}) = \begin{cases} \underline{z} & \text{for } \underline{k} = 1, \text{ and} \\ (1-\underline{z})^{\underline{k}-1}\underline{z} & \text{for } \underline{k} = 2, 3, \ldots . \end{cases} \tag{B3}$$

Total number of errors after the first success. Let ES be a random varibale that counts the total number of errors that occur after the first correct response. The distribution of ES is given by,

$$P(\underline{ES} = \underline{k}) = \begin{cases} \pi + (1-\pi)\theta < \{\underline{u} - [\underline{w}-(1-\underline{u})\underline{v}][\underline{u} - (1-\gamma)\rho]/[\gamma-(1-\underline{u})\underline{v}]\} \\ \{[\gamma-(1-\underline{u})\underline{v}]/[\underline{w}-(1-\underline{u})\underline{v}]\}\div(1-\underline{w})> + (1-\pi)\theta < \underline{u}\{1-[\gamma-(1-\underline{u})\underline{v}]/ \\ [\underline{w}-(1-\underline{u})\underline{v}]\}\div[1-(1-\underline{u})\underline{v}]\}> \quad \text{for } k = 0, \text{ and} \\ \\ (1-\pi)\theta\underline{u}[(1-\underline{u})\underline{v}]^{\underline{k}-1} < \{(1-\underline{u})(1-\underline{v}) - [\underline{w}-(1-\underline{u})\underline{v}][1-(1-\gamma)\rho]/ \\ [\gamma-(1-\underline{u})\underline{v}]\}\{[\gamma-(1-\underline{u})\underline{v}]/[\underline{w} - (1-\underline{u})\underline{v}]\}\div(1-\underline{w})> + (1-\pi)\theta\underline{u} \\ [(1-\underline{u})\underline{v}]^{\underline{k}-1} < (1-\underline{u})(1-\underline{v}) \{1-[\gamma-(1-\underline{u})\underline{v}]/[\underline{w}-(1-\underline{u})\underline{v}]\}\div[1-(1-\underline{u})\underline{v}]> \\ + (1-\pi)(1-\theta)\underline{u}[(1-\underline{u})\underline{v}]^{\underline{k}-1} \quad \text{for } \underline{k} = 1, 2, 3, \ldots . \end{cases} \tag{B4}$$

<u>States U and P</u>

Two statistics that provide information about the model's ability to account for the learning process as a whole are total errors and trial of last error.

<u>Total errors</u>. Let <u>TE</u> be a random variable that counts the total number of errors in a protocol. The distribution of <u>TE</u> is given by,

$$P(\underline{TE} = \underline{k}) = \begin{cases} \pi \quad \text{for } \underline{k} = 0, \text{ and} \\ \\ (1-\pi)[<1-\theta\{[\gamma-(1-\underline{u})\underline{v}]/[\underline{w}-(1-\underline{u})\underline{v}]\}\{[\underline{u}-(1-\gamma)\rho]/[\gamma-(1-\underline{u})\underline{v}]\} \\ \left[\dfrac{\underline{w}-(1-\underline{u})\underline{v}}{\underline{w}-(1-\underline{u})}\right]> \underline{u}(1-\underline{u})^{\underline{k}-1} + \\ \theta\{[\gamma-(1-\underline{u})\underline{v}]/[\underline{w}-(1-\underline{u})\underline{v}]\}\{[\underline{u}-(1-\gamma)\rho]/[\gamma-(1-\underline{u})\underline{v}]\} \\ \left[\dfrac{\underline{w}-(1-\underline{u})\underline{v}}{\underline{w}-(1-\underline{u})}\right] (1-\underline{w})\underline{w}^{\underline{k}-1}] \quad \text{for } \underline{k} = 1, 2, 3, \ldots . \end{cases} \tag{B5}$$

<u>Trial of last error</u>. Let <u>TLE</u> be a random variable that counts the trial number of the last error in each protocol. The distribution of <u>TLE</u> is given by,

$$P(\underline{TLE} = \underline{k}) = \begin{cases} \pi \quad \text{for } \underline{k} = 0, \text{ and} \\[1em] (1-\pi)\theta\langle \underline{u}-[\underline{w}-(1-\underline{u})\underline{v}]\{[\underline{u}-(1-\gamma)\rho]/[\gamma-(1-\underline{u})\underline{v}]\}\rangle\{[\gamma-(1-\underline{u})\underline{v}]/ \\ [\underline{w}-(1-\underline{u})\underline{v}]\}\underline{w}^{\underline{k}-1} + \underline{u}\{1-[\gamma-(1-\underline{u})\underline{v}]/[\underline{w}-(1-\underline{u})\underline{v}]\}[(1-\underline{u})\underline{v}]^{\underline{k}-1} \\ + (1-\pi)\theta\underline{uv}\ \{\sum_{\underline{i}=0}^{\underline{k}-2} (1-\underline{uv})^{\underline{k}-2-\underline{i}}\langle(1-\underline{u})(1-\underline{v})-[\underline{w}-(1-\underline{u})\underline{v}] \\ \{1-[\underline{u}-(1-\gamma)\rho]/[\gamma-(1-\underline{u})\underline{v}]\}\rangle\{[\gamma-(1-\underline{u})\underline{v}]/[\underline{w}-(1-\underline{u})\underline{v}]\}\underline{w}^{\underline{k}-1} \\ + (1-\underline{u})(1-\underline{v})\{1-[\gamma-(1-\underline{u})\underline{v}]/[\underline{w}-(1-\underline{u})\underline{v}]\}[(1-\underline{u})\underline{v}]^{\underline{k}-1}\} \\ \text{for } \underline{k} = 1, 2, 3, \ldots . \end{cases} \tag{B6}$$

By inserting the maximum-likelihood estimates of the relevant parameters, the null hypothesis of a statistically tolerable correspondence between the observed and predicted distributions can be assessed. One problem with using the chi-square test of goodness of fit is that the exact degrees of freedom for the asymtotic distributions of the chi-square are not known for these statistics. While Greeno (1968) has shown that a chi-square test can be approximated by constructing a range of critical values that bracket the true value, an alternative solution is to use a completely nonparametric test. As in previous research (e.g., Humphreys & Greeno, 1970), the test that was used here to assess goodness of fit was the Kolmogorov-Smirnov test.

APPENDIX C

Identification of the Theoretical Parameters of the Two-Stage Model

As mentioned earlier, only one of the theoretical parameters of Equation A1, namely, $\underline{a}$, is identifiable. In order to obtain numerical values of the remaining theoretical parameters, an identifying restriction must be located so that the system of 8 equations (Equations A4-A11) that maps the theoretical space onto the observable space can be reduced from 8 known (identifiable) parameters in 10 unknown (theoretical) parameters to a reduced system of 8 knowns and 8 unknowns. The identifying restriction that was found to be valid for all list conditions in this study was $\underline{r} = \underline{c} = 0$, and $1-\underline{a}' > \underline{a}'\underline{b}'(1-\underline{a}'\underline{b}')$. By inserting this restriction that Equations A4-A11, and performing the appropriate algebraic manipulations, the following definitions of the theoretical parameters of Equation A1, in terms of the identifiable parameters of Equation A2, were obtained.

$$\underline{t} = \pi \; ; \tag{C1}$$

$$\underline{s} = (1-\pi)\theta \; ; \tag{C2}$$

$$\underline{a}' = 1-(1-\pi)\theta \; ; \tag{C3}$$

$$\underline{a} = 1-\underline{w} \; ; \tag{C4}$$

$$\underline{b}' = \frac{\pi}{[1-(1-\pi)\theta]} \tag{C5}$$

$$\underline{b} = \underline{p}\left[1 - \frac{\gamma - \underline{w}}{1 - \underline{w}}\right] ; \tag{C6}$$

$$\underline{d} = \underline{u} \; ; \tag{C7}$$

$$\underline{g} = 1 - \underline{v} \; ; \tag{C8}$$

$$\underline{h} = 1 - \underline{z} \; ; \tag{C9}$$

$$\underline{e} = \frac{\rho - \underline{b}}{p(1-\underline{b})} \; . \tag{C10}$$

APPENDIX D

Associative Clusters used in Experiments 1 and 2.

Experiment 1: Unrelated clusters

Typical cue-Typical targets

boat	dog: spoon
roof	apple: sofa
pants	diamond: sword
screwdriver	hail: arm
gun	robin: car
foot	pea: train
bed	trout: socks
fork	elm: door
horse	nail: oak
pear	window: shark
opal	blouse: corn
snow	pan: eagle
sparrow	bomb: tornado
bean	head: emerald
herring	table: grape
maple	saw: lion

Typical cue-Atypical targets

boat	beaver: mirror
roof	raspberry: tooth
pants	crystal: duck
screwdriver	boots: rocker
gun	chicken: jeep
foot	rice: hatchet
bed	crab: haze
fork	beech: missile
horse	cellar: willow
pear	hallway: sardine
opal	girdle: skates
snow	board: cucumber
sparrow	dagger: frost
bean	skin: coral
herring	toaster: skunk
maple	saucer: date

Atypical cue-Typical targets

whip	robin: car
bar	trout: socks
waist	pea: train
kettle	elm: door
turtle	nail: oak
rhinestone	blouse: corn
olive	window: shark
drizzle	pan: eagle
garlic	head: emerald
mahogany	saw: lion
owl	bomb: tornado
minnow	table: grape
closet	apple: sofa
tape	hail: arm
submarine	dog: spoon
vest	diamond: sword

Atypical cue-Atypical targets

whip	chicken: jeep
bar	crab: haze
waist	rice: hatchet
kettle	beech: missile
turtle	cellar: willow
rhinestone	girdle: skates
olive	hallway: sardine
drizzle	board: cucumber
garlic	skin: coral
mahogany	saucer: date
owl	dagger: frost
minnow	toaster: skunk
closet	raspberry: totth
tape	boots: rocker
submarine	beaver: mirror
vest	crystal: duck

Experiment 2: Related clusters

Typical cue-Typical targets

boat	car: train
roof	window: door
pants	blouse: socks
screwdriver	saw: nail
gun	bomb: sword
foot	head: arm
bed	table: sofa
fork	pan: spoon
horse	dog: lion
pear	apple: grape
opal	diamond: emerald
snow	hail: tornado
sparrow	robin: eagle
bean	pea: corn
herring	trout: shark
maple	elm: oak

Typical cue-Atypical targets

boat	jeep: skates
roof	hallway: cellar
pants	girdle: boots
screwdriver	board: hatchet
gun	dagger: missile
foot	skin: tooth
bed	mirror: rocker
fork	toaster: saucer
horse	beaver: skunk
pear	raspberry: date
opal	crystal: coral
snow	haze: frost
sparrow	chicken: duck
bean	rice: cucumber
herring	crab: sardine
maple	beech: willow

Atypical cue-Typical targets

whip	bomb: sword
bar	table: sofa
waist	head: arm
kettle	pan: spoon
turtle	dog: lion
rhinestone	diamond: emerald
olive	apple: grape
drizzle	hail: tornado
garlic	pea: corn
mahogany	elm: oak
owl	robin: eagle
minnow	trout: shark
closet	window: door
tape	saw: nail
submarine	car: train
vest	blouse: socks

Atypical cue-Atypical targets

whip	dagger: missile
bar	mirror: rocker
waist	skin: tooth
kettle	toaster: saucer
turtle	beaver: skunk
rhinestone	crystal: coral
olive	raspberry: date
drizzle	haze: frost
garlic	rice: cucumber
mahogany	beech: willow
owl	chicken: duck
minnow	crab: sardine
closet	hallway: cellar
tape	board: hatchet
submarine	jeep: skates
vest	girdle: boots

END

11·01·83

FIN